CONTENT 2014

The Sixth International Conference on Creative Content Technologies

ISBN: 978-1-61208-342-1

May 25 - 29, 2014

Venice, Italy

CONTENT 2014 Editors

Hans-Werner Sehring, T-Systems Multimedia Solutions GmbH, Germany

René Berndt, Fraunhofer Austria Research GmbH, Austria

CONTENT 2014

Foreword

The Sixth International Conference on Creative Content Technologies (CONTENT 2014), held between May 25-29, 2014 in Venice, Italy, targeted advanced concepts, solutions and applications in producing, transmitting and managing various forms of content and their combination. Multi-cast and uni-cast content distribution, content localization, on-demand or following customer profiles are common challenges for content producers and distributors. Special processing challenges occur when dealing with social, graphic content, animation, speech, voice, image, audio, data, or image contents. Advanced producing and managing mechanisms and methodologies are now embedded in current and soon-to-be solutions.

We take here the opportunity to warmly thank all the members of the CONTENT 2014 Technical Program Committee, as well as all of the reviewers. The creation of such a high quality conference program would not have been possible without their involvement. We also kindly thank all the authors who dedicated much of their time and efforts to contribute to CONTENT 2014. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

Also, this event could not have been a reality without the support of many individuals, organizations, and sponsors. We are grateful to the members of the CONTENT 2014 organizing committee for their help in handling the logistics and for their work to make this professional meeting a success.

We hope that CONTENT 2014 was a successful international forum for the exchange of ideas and results between academia and industry and for the promotion of progress in the area of creative content technologies.

We are convinced that the participants found the event useful and communications very open. We hope that Venice, Italy, provided a pleasant environment during the conference and everyone saved some time to enjoy the charm of the city.

CONTENT 2014 Chairs:

Raouf Hamzaoui, De Montfort University - Leicester, UK
Jalel Ben-Othman, Université de Versailles, France
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain
Wolfgang Fohl, Hamburg University of Applied Sciences, Germany
Zhou Su, Waseda University, Japan
Ajith Abraham, Machine Intelligence Research Labs (MIR Labs), USA
Hans-Werner Sehring, T-Systems Multimedia Solutions GmbH, Germany
René Berndt, Fraunhofer Austria Research GmbH, Austria
Lorena Parra, Universidad Politécnica de Valencia, Spain
Samuel Kosolapov, Braude Academic College of Engineering, Israel
Wilawan Inchamnan, Queensland University of Technology, Australia
Javier Quevedo-Fernandez, Eindhoven University of Technology, The Netherlands

CONTENT 2014

Committee

CONTENT Advisory Chairs

Raouf Hamzaoui, De Montfort University - Leicester, UK
Jalel Ben-Othman, Université de Versailles, France
Jaime Lloret Mauri, Polytechnic University of Valencia, Spain
Wolfgang Fohl, Hamburg University of Applied Sciences, Germany
Zhou Su, Waseda University, Japan

CONTENT Industry/Research Chairs

Ajith Abraham, Machine Intelligence Research Labs (MIR Labs), USA
Hans-Werner Sehring, T-Systems Multimedia Solutions GmbH, Germany
René Berndt, Fraunhofer Austria Research GmbH, Austria

CONTENT Publicity Chairs

Lorena Parra, Universidad Politécnica de Valencia, Spain
Samuel Kosolapov, Braude Academic College of Engineering, Israel
Wilawan Inchamnan, Queensland University of Technology, Australia
Javier Quevedo-Fernandez, Eindhoven University of Technology, The Netherlands

CONTENT 2014 Technical Program Committee

Marios C. Angelides, Brunel University - Uxbridge, UK
Kambiz Badie, Research Institute for ICT & University of Tehran, Iran
David Banks, University of Tennessee, USA
Christos Bouras, University of Patras and Computer Technology Institute & Press «Diophantus», Greece
Yiwei Cao, RWTH Aachen University, Germany
Wojciech Cellary, Poznan University of Economics, Poland
Lijun Chang, University of New South Wales, Australia
Savvas A. Chatzichristofis, Democritus University of Thrace, Greece
Chi-Hua Chen, National Chiao Tung University, Taiwan, R.O.C.
Octavian Ciobanu, "Gr.T. Popa" University of Medicine and Pharmacy – Iasi, Romania
Raffaele De Amicis, Fondazione Graphitech - Trento, Italy
Rafael del Vado Vírseda, Universidad Complutense de Madrid, Spain
Marco di Benedetto, ISTI - National Research Council (CNR), Italy
Eva Eggeling, Fraunhofer Austria Research GmbH, Austria
Klemens Ehret, University of Applied Sciences Ravensburg-Weingarten, Germany
Wolfgang Fohl, Hamburg University of Applied Sciences, Germany
Antonio Javier García Sánchez, Technical University of Cartagena, Spain
Afzal Godil, National Institute of Standards and Technology, USA
Patrick Gros, INRIA Rennes - Campus de Beaulieu, France
Raouf Hamzaoui, De Montfort University - Leicester, UK

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

Effects of Speaking Rate on Initial and Final Duration Structure in Mandarin Chinese

Wen-Hsing Lai
Department of Computer and Communication Engineering
National Kaohsiung First University of Science and Technology
Kaohsiung, Taiwan
e-mail: lwh@nkfust.edu.tw

***Abstract*—An Expectation-Maximization (EM) modeling and a speech corpus with fast, median, and slow speaking rate are applied to explore the effect of speaking rate on segmental duration structure of *Initial*, *Final*, and syllable in Mandarin Chinese. Experimental results showed that the variance of duration was greatly reduced after eliminating effects from additive factors by EM algorithm. By excluding the interference of acoustical factors, the relationship between syllable duration and the structure of *Initial* and *Final* durations for different speaking rate is observed. The result shows that for same syllable duration, the ratio of *Final* to *Initial* becomes larger when the speaking rate becomes faster. Besides, the ratio, generally, becomes larger as the syllable becomes longer. However, for extremely short syllable about less than 100 ms in fast speed, the ratio becomes large, and in syllable duration longer than about 350 ms in median and slow speed, the ratio becomes almost a constant.**

Keywords- speaking rate; duration; Mandarin Chinese

I. Introduction

Speaking rate is one of the most important factors in spontaneous speech systems, because variability in speaking rate may be often observed in spontaneous speech than in read speech. Studies have shown that the acoustic properties corresponding to phonetic segments of speech are influenced by variability of speaking rate. For example, spectral patterns will be changed and formant positions will be shifted [1]-[4] and the intelligibility and comprehension will be influenced [5]-[8]. Furthermore, changes in speech rate have effects on prosody, like the overall level and range of fundamental frequency (*F*0) and durations. While changing speaking rates, it also causes variations in prosodic phrasing, such as prosodic boundaries. For example, the research of Tseng find that duration adjustment is made systematically at each prosody level during speech production and examining speech rate in relation to prosody units is a significant first step to understanding temporal organization of speech flow [9].

Because of its importance, speaking rate has been put into considerations in many speech application, for instance, speech recognition [10]-[16], emotion classification [17], and Text-To-Speech system [18]. Previous studies revealed that the mismatch between speaking rates of training and test data of speech recognition system will degrade the system performance, therefore, some researches focus on solving the problem of performance degradation caused by speaking rate variability [10]-[16]. Further, different emotional dispositions of a person are strongly expressed in his/her speaking rate [17], therefore, speaking rate also has significant influence in emotion classification. Hence, speaking rate estimation [19] has become an important job. In addition, speaking rate-controlled prosody is also critical for Text-To-Speech system [18].

Though there are some studies about the effect of speaking rate for vowel duration [20] and prosody units [9], there are few studies about the initial and final structure change under various speaking rate. Therefore, our objective of this paper is to find out the change of initial and final structure under various speaking rate to further understand the temporal organization of speech flow, so it could be applied to speech related application.

However, it is difficult to get an obvious pattern from observing the original duration directly or to incorporate qualitative findings into a quantitative model, and there has been rather few prosodic model devoted to investigating detailed effects of speech rate modification on the realization of individual pitch accents, duration, intonation, and prosodic structures. Hence, an Expectation-Maximization (EM) modeling [21] and a speech corpus with fast, median, and slow speaking rate are applied to explore the effect of speaking rate on segmental duration structure in Mandarin Chinese in this article. The achievement will be useful for improving the quality of speech synthesis and the recognition rate of speech recognition.

The paper is organized as follows. In Section II, the EM analysis algorithm, including the factors which have impact on durations, the syllable duration modeling, and the extension to *Initial* and *Final* Duration Modeling, is shown. Section III describes the experimental results. Conclusions are given in the last section.

II. EM Analysis Algorithm

A. Factors

In naturally spoken Mandarin Chinese, duration varies considerably depending on various linguistic and nonlinguistic factors [22]. Mandarin Chinese is a tonal and syllable-based language. Each character is pronounced as a syllable, the basic pronunciation unit. There exist only about 1300 phonetically distinguishable syllables comprising all legal combinations of approximately 411 base-syllables and five tones. Mandarin base-syllables have very regular phonetic structure. Each base-syllable is of the form (C)(C)V(N), where C is a consonant, V is a vowel, and N is a nasal consonant (the symbols between parentheses signal optionality). So, base-syllables comprise one to four phonemes. Generally speaking, syllable duration increases as the number of constituent phonemes increases. Syllables with single vowels are shortest. Syllables with stop initials or no initials, and without nasal endings are pronounced shorter. Syllables with fricative initials and with nasal endings are longer. Therefore, duration is seriously influenced by the phonetic structure of base-syllables, and base-syllable is listed as one of the impact factors.

Tonality of a syllable is characterized by its pitch contour shape, loudness and duration. For example, syllables with Tone 5 (Neutral Tone) are always pronounced much shorter. Therefore, tonality is also considered as a factor which has impact on duration.

To prevent the speed variation in different utterance sentence in recording process for specific speed rate, utterance is also included as a factor for normalization.

Aside from the acoustic factors mentioned above, including lexical tone, base-syllable, and utterance, other high-level linguistic components, such as word-level and syntactic-level factors, like the boundary index, position in word, length of word factors used in [22], also seriously influence the duration of an utterance. In the model, the prosodic state, as a substitute for high-level linguistic information, is used to indicate the state in prosodic word or prosodic phrases, for example, to indicate the possible prosodic word or phrase boundaries, or the notions of prominence. Therefore, the prosodic state is used to account for the influence of all high-level linguistic features. There are two advantages of using the prosodic state to replace high-level linguistic features. Firstly, duration information is a prosodic feature, so the variation of the duration should better match the prosodic phrase structure than the syntactic phrase structure. Secondly, as mentioned above, some unsolved problems, such as the ambiguity of word-segmentation and word-chunking in Mandarin Chinese and the difficulty of performing automatic syntactic analysis on unlimited natural texts, can be avoided in the current duration modeling approach. This prevents us from using improper or incomplete high-level linguistic information. By doing so, the modeling of duration can simply consider the effects of prosodic state and acoustical factors, like tone, utterance and base-syllable factors. Due to the fact that the prosodic state is not explicitly given, it has been treated as a hidden variable in the EM algorithm. The number of prosodic states is set as 16 in our modeling. A by-product of the EM algorithm is the determination of the hidden prosodic states of all the units in the training set. This is an additional advantage. From the sequence of prosodic states, some high-level linguistic phenomenon could be observed, like the possible prosodic phrase boundaries.

In sum, four major affecting factors including tone, base-syllable, utterance, and prosodic state are considered.

B. Syllable Duration Modeling

By considering the factors in Section II-A, an additive duration model can be expressed by

$$Z_n = X_n + \gamma_{t_n} + \gamma_{y_n} + \gamma_{j_n} + \gamma_{l_n} , \qquad (1)$$

where Z_n and X_n are, respectively, the observed duration and the normalized (residual) duration of the *n*th syllable. X_n is considered as the residual duration after excluding all the impact from factors and is modeled as a normal distribution with mean μ and variance ν . γ_{t_n} , γ_{y_n} , γ_{j_n} , and γ_{ln} are the impact value of the lexical tone, prosodic state, base-syllable, and utterance identification number factor of the *n*th syllable, indicated by t_n , y_n , j_n and l_n .

To illustrate the EM algorithm, an auxiliary function is defined in the expectation step as

$$Q(\bar{\lambda},\lambda) = \sum_{n=1}^{N}\sum_{y_n=1}^{Y} p(y_n \mid Z_n,\bar{\lambda}) \log p(Z_n,y_n \mid \lambda) , \qquad (2)$$

where *N* is the total number of training samples; *Y* is the total number of prosodic states; $\lambda = \{\mu,\nu,\gamma_t,\gamma_y,\gamma_j,\gamma_l\}$ is the set of parameters to be estimated, and λ and $\bar{\lambda}$ are, respectively, the updated and old parameter sets. γ_t , γ_y , γ_j , and γ_l represent the impact value of all the lexical tone, prosodic state, base-syllable, and utterance identification number factor. For example, the possible t_n , the lexical tone of the *n*th syllable, is 1 to 5, therefore, γ_t represent γ_1 , γ_2 , γ_3 , γ_4 , and γ_5 .

$$p(Z_n, y_n \mid \lambda) = N(Z_n; \mu + \sum_p \gamma_{p_{\mathrm{n}}}, \nu) , \qquad (3)$$

where $N(Z;\mu,\nu)$ is a normal distribution of *Z* with mean μ and variance ν . $p(y_n \mid Z_n,\bar{\lambda})$ can be represented as

$$p(y_n \mid Z_n,\bar{\lambda}) = \frac{p(Z_n, y_n \mid \bar{\lambda})}{\sum_{y'_n=1}^{Y} p(Z_n, y'_n \mid \bar{\lambda})} . \qquad (4)$$

To cure the drawback of the non-uniqueness of the solution because of the use of additive factors, the optimization procedure in the Maximization step (M-step) is modified to a constrained optimization via introducing a global duration constraint. The auxiliary function then changes to

$$Q(\bar{\lambda},\lambda)=\sum_{n=1}^{N}\sum_{y_n=1}^{Y}p(y_n\,|\,Z_n,\bar{\lambda})\log p(Z_n,y_n\,|\,\lambda) \\ +\eta(\sum_{n=1}^{N}(\mu+\sum_{p}\gamma_{p_n})-N\mu_z) \quad , \qquad (5)$$

where μ_z is the average of Z_n and η is a Lagrange multiplier [23]. The constrained optimization is finally solved by the Newton-Raphson method [23].

Initializations of the parameters in λ are done by estimating each parameter independently. Then, iterative sequential optimizations of the parameters in λ are performed in the M-step. Iterations are continued until a convergence is reached. The prosodic state can finally be assigned by

$$y_n^* = \arg\max_{y_n} p(y_n\,|\,Z_n,\lambda)\;. \qquad (6)$$

C. Extension to Initial and Final Duration Modeling

Each Mandarin syllable is composed of an optional consonant *Initial* and a *Final*. The *Final* comprises an optional medial, a vowel nucleus and an optional nasal ending. To exploit the relationship between the syllable duration and its component *Initial* and *Final* durations in different speaking rate, the above syllable duration modeling is extended to the duration modeling of *Initial* and *Final*. There are two approaches. One approach is to keep the prosodic states of the three models independent because the optimal prosodic states of both *Initial* and *Final* duration models may not match with those of the syllable duration model. The mismatch may results from the inconsistency in the effect of linguistic features on the *Initial* duration and on the *Final* duration. A previous study [24] found that consonant-lengthening can happen at all initial positions especially at the beginning of a word, while vowel-lengthening can occur only at phrasal final. The other approach is to share the same prosodic states so the relationship between the impact value of prosodic states of syllable and those of *Final* and *Initial* can be observed more conveniently. For the first approach, *Initial* and *Final* models could adopt the similar method as syllable. For the sharing model, an additional constraint is set in *Initial* and *Final* models to let their prosodic states the same as in syllable model. The training algorithm of *Initial* and *Final* models is then modified to an ML (Maximum Likelihood) one with all prosodic states being predetermined by the training procedure of syllable model.

III. Experimental Results

The corpus is recorded in fast, median, and slow speed by a professional female announcer in reading style using WaveSurfer software on personal computer. The median speed was recorded first. The material contains 359 short paragraphs including news, blogs and text books of elementary school. There are totally 44934 syllables. Averagely, every sentence contains 10.37 syllables. The sentence length ranges from 80 to 272 syllables. The sampling rate is 20 kHz and the file format is 16 bit PCM. The pronunciations have been labeled. The boundaries of syllable, *Initial* and *Final* have also been marked by automatic segmentation based on Hidden Markov Model ToolKit (HTK) [25], and then corrected manually.

Table 1 shows the duration mean, standard deviation and the ratio of standard deviation to mean of syllable, *Initial* and *Final* in fast, median, and slow speed. The experiment of *Initial* was done without considering the null *Initial* and the very short *Initial*s of {b, d, g} which are generally difficult to be segmented accurately. As shown in Table 1(a) and (b), the duration mean of syllable, *Initial* and *Final* lengthen and the standard deviation become larger as the speed slows down. Besides, from Table 1(c), the ratio of standard deviation to mean of fast speed is the largest. That is, the relative variation is the greatest in high speed.

The normal distribution assumption is then checked. Take syllable durations in slow speed as an example. Fig. 1(a) shows the histogram with a fitted normal distribution and Fig. 1(b) shows a normal Q-Q plot with an RMA (Reduced Major Axis) regression line, together with the Probability Plot Correlation Coefficient (PPCC) equals 0.9967. (Basically, a normal distribution will plot on a straight line.) Besides, Shapiro-Wilk normality test returns a test statistic W = 0.9934, ($0 < W \leqq 1$, W is small for non-normal samples). Jarque-Bera normality test returns JB = 870.1 and chi-square normality test returns value 200.01. The p values of the three tests are much smaller than 0.05. From the above observations, except some outliers (those samples with much longer and shorter durations), most samples actually fit the normal distribution. To make the model simple, the assumption of Gaussian distribution is still adopted in this study. A mixture Gaussian distribution may fit better and could be put as a future study.

Table 2 shows the mean, standard deviation, and RMSE (Root Mean Square Error) of the normalized duration of syllable, *Final*, and *Initial* in fast, median, and slow speed in EM modeling. The *Final* and *Initial* models used in this experiment are not sharing prosodic states with syllable model. After excluding the impact of factors by EM modeling, the standard deviation of the normalized duration in Table 2(b) greatly reduced compared with the original standard deviation in Table 1(b), while the mean of the normalized duration in Table 2(a) is almost the same with the mean in Table 1(a). Therefore, the EM modeling can successfully exclude the impact of factors. The RMSE of prediction duration by additive model is shown in Table 2(c). From Table 1(b), 2(b), and 2(c), though the original syllable standard deviation of high speed speaking rate is higher than the deviation of *Final*, the normalized syllable standard deviation of high speed speaking rate becomes lower than the deviation of *Final*, and in prediction stage, the RMSE of syllable prediction is lower than the RMSE of *Final*. The relatively high deviation of the normalized *Final* duration and RMSE show that the *Final* duration in fast speed is more difficult to model than syllable duration.

At last, the relationship between syllable duration and the structure of *Initial* and *Final* durations after excluding the

TABLE I. THE DURATION (A) MEAN (UNIT: MS), (B) STANDARD DEVIATION (UNIT: MS) AND (C) RATIO OF STANDARD DEVIATION/MEAN OF SYLLABLE, *INITIAL* AND *FINAL* IN FAST, MEDIAN, AND SLOW SPEED.

(a)

	Fast	Median	Slow
Syllable	185.253	245.541	271.494
Final	135.001	176.731	195.280
Initial	73.265	97.643	107.517

(b)

	Fast	Median	Slow
Syllable	67.683	77.259	83.390
Final	62.291	67.527	75.388
Initial	37.679	44.210	48.407

(c)

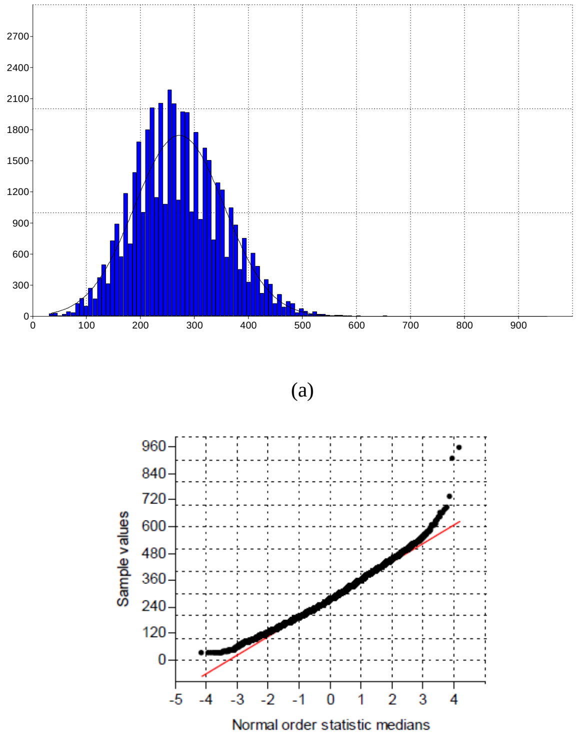
	Fast	Median	Slow
Syllable	0.365	0.315	0.307
Final	0.461	0.382	0.386
Initial	0.514	0.453	0.450

interference from acoustical factors for different speaking rate is examined. The impact value of prosodic states plus the mean of duration is taken as the duration excluding the impact from acoustic factors. Specifically, the ratio of $(\gamma_y^f+\mu_f)/(\gamma_y^i+\mu_i)$ versus $(\gamma_y+\mu)$ is observed. γ_y^f, γ_y^i, and γ_y are the impact value of prosodic states of *Final*, *Initial* and syllable duration models. μ_f, μ_i, and μ are the mean of *Final*, *Initial* and syllable durations. $(\gamma_y^f+\mu^f)/(\gamma_y^i+\mu^i)$ can be considered as the duration component ratio of *Final* to *Initial* in syllable structure. For easy comparing, the *Final* and *Initial* models used in this experiment are sharing prosodic states with syllable model.

Fig. 2 displays the figure of $(\gamma_y^f+\mu_f)/(\gamma_y^i+\mu_i)$ versus $(\gamma_y+\mu)$, or the ratio of *Final* to *Initial* versus the syllable duration after excluding the interference from acoustical factors. The vertical axis is $(\gamma_y^f+\mu_f)/(\gamma_y^i+\mu_i)$ and the horizontal axis is $(\gamma_y+\mu)$. From Fig. 2, it is easy to see that for the same syllable duration, the duration ratio of *Final* to *Initial* of fast speaking rate is highest. It is followed by the ratio of median rate, and the ratio of slow rate is lowest. That is, the ratio of *Final* to *Initial,* generally, becomes larger as the speaking rate increases. It may be because in fast speed, the pronunciation is more relaxed and *Final* dominates. Besides, generally, the ratio becomes larger as the syllable becomes longer. But, for extremely short syllable in fast speed, about less than 100 ms, the ratio becomes large. Besides, in syllable duration larger than about 350 ms in median and slow speed, the value of the ratio gets almost saturated and becomes almost a constant.

Our objective of this modeling is to find out the change of initial and final structure under various speaking rate. However, it is difficult to get an obvious pattern from the original observed duration. Observing the results of our experiment, the value of acoustic factors including lexical tone, base-syllable, and utterance, did not show particular different pattern among different speaking rate. Therefore, we assume that speaking rate does not have big impact on the acoustic factors including lexical tone, base-syllable, and utterance in our experiment, and we observed the change of initial and final structure under various speaking rate as in Fig.2. After excluding the interference of acoustic factors, it is easy to find out that for the same syllable duration, when we increase the speaking rate, the duration ratio of *Final* to *Initial* becomes larger.

At last, speaker factor may be also important for speaking rate. Since our experiment is based on a corpus recorded by a single professional speaker, the impact of speaker is not included in our modeling, therefore, the further study of speaker factor will be put as our future work.



(a)

(b)

Figure 1. The (a) histogram (unit of horizontal axis: ms) (b) normal Q-Q plot with an RMA regression line (unit of vertical axis: ms) of the observed syllable durations in slow speed.

TABLE II. THE (A) MEAN, (B) STANDARD DEVIATION, AND (C) RMSE OF THE NORMALIZED DURATION OF SYLLABLE, *FINAL*, AND *INITIAL* IN FAST, MEDIAN, AND SLOW SPEED (UNIT: MS).

(a)

	Fast	Median	Slow
Syllable	183.291	242.985	268.032
Final	133.847	175.115	192.651
Initial	73.539	97.956	107.382

(b)

	Fast	Median	Slow
Syllable	8.928	11.602	11.596
Final	11.660	10.544	11.447
Initial	5.204	6.489	7.051

(c)

	Fast	Median	Slow
Syllable	8.933	11.608	11.600
Final	11.677	10.569	11.450
Initial	5.233	6.520	7.087

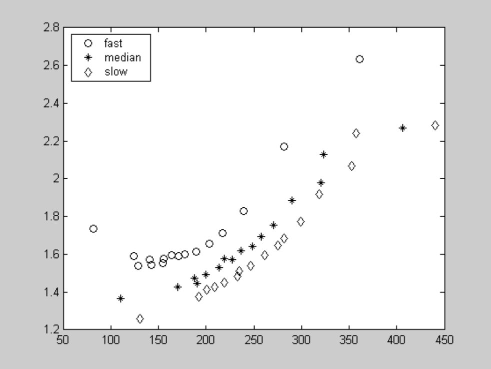


Figure 2. The ratio of *Final* to *Initial* versus syllable duration (unit: ms) after excluding acoustical factors.

IV. CONCLUSIONS

In this paper, the duration variation was studied and duration models are built for syllable, *Initial* and *Final* in different speaking rate for Mandarin Chinese. An EM algorithm is applied to syllable duration modeling. Extensions of the syllable duration modeling method are also performed on *Initial* and *Final*. From the experimental results, the impact of factors on syllable, *Initial* and *Final* duration in different speaking rate are explored. By observing the relationship between syllable duration and the structure of *Initial* and *Final* durations after excluding the interference from acoustical factors for different speaking rate, an important conclusion is that for the same syllable duration, the duration ratio of *Final* to *Initial* becomes larger as the speaking rate increases. In addition, the ratio basically becomes larger as the syllable becomes longer. But for extremely short syllable in fast speed, the ratio becomes large; in syllable duration larger than about 350 ms in median and slow speed, the ratio becomes almost saturated.

ACKNOWLEDGMENT

This work was supported by NSC, Taiwan under Contract NSC 99-2221-E-327-044. The author thanks Dept. of Electrical and Computer Engineering, National Chiao Tung University, Taiwan for supplying the speech databases.

REFERENCES

[1] M. Pitermann, “Effect of Speaking Rate and Contrastive Stress on Formant Dynamics and Vowel Perception,” The Journal of the Acoustical Society of America, vol. 107, no. 6, Jul. 2000, pp. 3425-3437, doi:10.1121/1.429413.

[2] T. Gay, “Effect of Speaking Rate on Vowel Formant Movements,” The Journal of the Acoustical Society of America, vol. 63, no. 1, Feb. 1978, pp. 223-230, doi:10.1121/1.381717.

[3] G. Weismer and J. Berry, “Effects of Speaking Rate on Second Formant Trajectories of Selected Vocalic Nuclei,” The Journal of the Acoustical Society of America, vol. 113, no. 6, Jul. 2003, pp. 3362-3378, doi:10.1121/1.1572142.

[4] D. O’Shaughnessy, “The Effects of Speaking Rate on Formant Transitions in French Synthesis-by-Rule,” Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 86), vol. 11, Apr. 1986, pp. 2027-2030, doi: 10.1109/ICASSP.1986.1168797.

[5] A. H. S. Chan and P. S. K. Lee, “Intelligibility and Preferred Rate of Chinese Speaking,” International Journal of Industrial Ergonomics, vol. 35, no. 3, Jan. 2005, pp. 217-228, doi:10.1016/j.ergon.2004.09.001.

[6] J. C. Krause and L. D. Braida, “Investigating Alternative Forms of Clear Speech: The Effects of Speaking Rate and Speaking Mode on Intelligibility,” The Journal of the Acoustical Society of America, vol. 112, no. 5, pt. 1, Nov. 2002, pp. 2165-2172, doi: 10.1121/1.1509432.

[7] C. Jones, L. Berry, and C. Stevens, “Synthesized Speech Intelligibility and Persuasion: Speech Rate and Non-native Listeners,” Computer Speech and Language, vol. 21, 2007, pp. 641-651, doi:10.1016/j.csl.2007.03.001.

[8] S. Liu and F. G. Zeng, “Temporal Properties in Clear Speech Perception,” The Journal of the Acoustical Society of America, vol. 120, no. 1, Jul. 2006, pp. 424-432, doi: 10.1121/1.2208427.

[9] C. Y. Tseng and Y. L. Lee, “Speech rate and prosody units: Evidence of interaction from Mandarin Chinese,” Proceedings of the International Conference on Speech Prosody, Mar. 2004, pp. 251-254.

[10] E. Fosler-Lussier and N. Morgan, “Effects of Speaking Rate and Word Frequency on Pronunciations in Convertional Speech,” Speech Communication, vol. 29, no. 2-4, Jan. 1999, pp. 137-158, doi:10.1016/S0167-6393(99)00035-7.

[11] T. Shinozaki and S. Furui, “Hidden Mode HMM using Bayesian Network for Modeling Speaking Rate Fluctuation,”

Proc. IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU03), Jan. 2003, pp. 417-422, doi:10.1109/ASRU.2003.1318477.

[12] H. Nanjo and T. Kawahara, “Language Model and Speaking Rate Adaptation for Spontaneous Presentation Speech Recognition,” IEEE Transactions on Speech and Audio Processing, vol. 12, issue 4, Jul. 2004, pp. 391-400, doi: 10.1109/TSA.2004.828641.

[13] M. S. Sommers, L. C. Nygaarda, and D. B. Pisoni, “Stimulus Variability and Spoken Word Recognition. I. Effects of Variability in Speaking Rate and Overall Amplitude,” The Journal of the Acoustical Society of America, vol. 96, no. 3, Sep. 1994, pp. 1314-1324, doi: 10.1121/1.411453.

[14] M. S. Sommers and J. Barcroft, “Stimulus Variability and the Phonetic Relevance Hypothesis: Effects of Variability in Speaking Style, Fundamental Frequency and Speaking Rate on Spoken Word Identification,” The Journal of the Acoustical Society of America, vol. 119, no. 4, Apr. 2006, pp. 2406-2416, doi: 10.1121/1.2171836.

[15] M. Radeau, J. Morais, P. Mousty, and P. Bertelson, “The Effect of Speaking Rate on the Role of the Uniqueness Point in Spoken Word Recognition,” Journal of Memory and Language, vol. 42, 2000, pp. 406-422, doi:10.1006/jmla.1999.2682.

[16] S. M. Ban and H. S. Kim, “Speaking Rate Dependent Multiple Acoustic Models Using Continuous Frame Rate Normalization,” Asia-Pacific Signal & Information Processing Association Annual Summit and Conference (APSIPA ASC), Dec. 2012, pp. 1-4.

[17] D. Philippou-Hubner, B. Vlasenko, R. Bock, and A. Wendemuth, “The Performance of the Speaking Rate Parameter in Emotion Recognition from Speech,” IEEE International Conference on Multimedia and Expo Workshops (ICMEW), July 2012, pp. 296-301, doi: 10.1109/ICMEW.2012.57.

[18] C. H. Hsieh, Y. R. Wang, C. Y. Chiang, and S. H. Chen, “A Speaking rate-Controlled Mandarin TTS System,” 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2013, pp. 6900-6904, doi: 10.1109/ICASSP.2013.6638999.

[19] Y. Wu, Q. H. He, and Y. X. Li, “Speaking Rate Estimation for Multi-Speakers,” International Conference on Audio, Language and Image Processing (ICALIP), July 2012, pp. 976-979, doi: 10.1109/ICALIP.2012.6376756.

[20] M. Wang, W. Shi, R. Huang, and Z. Xiong, “The Temporal Effect of Speaking Rate, Focus and Prosody in Chinese,” 8th International Symposium on Chinese Spoken Language Processing (ISCSLP), Dec. 2012, pp. 445-449, doi: 10.1109/ISCSLP.2012.6423481.

[21] S. H. Chen, W. H. Lai, and Y. R. Wang, “A new duration modeling approach for Mandarin speech,” IEEE Trans. on Speech and Audio Processing, vol.11, issue 4, July 2003, pp. 308-320, doi: 10.1109/TSA.2003.814377.

[22] M. Chu and Y. Feng, “Study on factors influencing durations of syllables in Mandarin,” EUROSPEECH, Sep. 2001, pp. 927-930.

[23] P. Wriggers, Computational Contact Mechanics, 2nd ed., Spring, 2006, pp. 336-337.

[24] C. Shih and B. Ao, “Duration Study for the Bell Laboratories Mandarin Text-to-Speech System,” in Progress in Speech Synthesis, J. P. H. van Santen, J. P. Olive, R. W. Sproat, and J. Hirschberg, Eds. Springer, 1997, pp. 383-399.

[25] S. J. Young, “The HTK Hidden Markov Model Toolkit: Design and Philosophy,” University of Cambridge, Department of Engineering, 1994, http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=BD1F4D060CB93D921A658DB57F3C1839?doi=10.1.1.17.8190&rep=rep1&type=pdf.

Adaptive Playout Control and Signal Reconstruction for Speech-Based Audio Convergence VoIP

Jun-Yong Lee
Department of Electronics Convergence Engineering
Kwangwoon University
Seoul, South-Korea
Jasonlee88@kw.ac.kr

Hyoung-Gook Kim
Department of Electronics Convergence Engineering
Kwangwoon University
Seoul, South-Korea
hkim@kw.ac.kr

***Abstract*—This paper proposes an adaptive playout control and signal reconstruction method for speech-based audio convergence VoIP. In adaptive playout control, the buffering time is minimized by way of playing out normally or compressing each packet according to accurate network jitter estimation. Also, linear prediction-based signal reconstruction recovers lost packets and minimizes boundary discontinuities between the good packets and the reconstructed packets. The proposed receiver-based enhancing method delivers high-quality voice and music service over IP networks.**

Keywords-playout control; signal reconstrution; jitter estimation

I. INTRODUCTION

The use of Voice over Internet Protocol (VoIP) for carrying real-time voice data over any IP network has significant impacts on the telecommunication industry.

However, a number of factors may affect the service quality of VoIP, such as packet loss, packet delay, and network delay variation (also known as "jitter"). To provide reliable services with satisfactory voice quality over IP networks, considerable efforts [1][2][3][4][5] have been made within different layers of current communication systems to reduce delay, smooth jitter, and recover loss.

Some techniques have been developed for concealing packet loss. In waveform substitution method, the missing frames are replaced by another already-received frame using pitch replication [4], [6] or pattern matching [7]. And the model of the previously received signal (eventually slightly modified) is used to generate the missing signal [8] in model extrapolation method.

Several VoIP playout buffer scheduling or timing recovery algorithms have been proposed. Pinto [9] presented a method that adjusts silence periods between signal spurts to improve voice interaction quality, while Liang [3] proposed adaptive playout-buffer schedulers that adjust the voice regions by introducing time-scale modification. Chi [10], Li [5], and Aragao [11] suggest a playout scheduling method based on modeling packet arrival times using K-Erlang distribution, a Gaussian model, Pareto distribution, etc. However, these methods are "packet-based" and decide whether to stretch or compress a packet once it is received. Florncio [12] used the "buffer-based" method, which decides whether to stretch or compress only when the audio playout device needs a frame.

According to enhancing VoIP speech quality, more and more smartphone users are taking advantage of mobile VoIP services. Recently, speech-based audio convergence VoIP codecs [13] that offer a high or nearly transparent quality while remaining compliant with tight conversational requirements (delay constraints in two-way communication) are recently emerging for the applications of the high-quality conferencing and VoIP telephony. Specially, ITU-T Recommendation G.729.1 is a scalable wideband speech and audio coding standard designed to facilitate a graceful and cost-effective evolution to high-quality wideband speech based audio communications in packet-switched networks.

In this paper, we focus on enhancing VoIP speech and music quality only at receiving portion of a mobile Internet phone. The important functionality to be implemented at the receiver is an adaptive playout control and signal reconstruction scheme consisting of concealment of lost packets based on the redundancy in neighboring packets, adaptive playout-buffer scheduling using active jitter estimation, and smooth interpolation between two signals in a transition region.

Our method has three important improvements: 1) using accurate jitter estimation our playout-buffer control makes it possible to trade-off the buffering time with the rate of packet loss; 2) our signal reconstruction based on recursive linear prediction analysis and synthesis (LPAS) alleviates the metallic artifacts that are often introduced during concealing packet loss; and 3) using linear prediction (LP) based smooth interpolation between the two signals in a transition region, we improve VoIP speech and voice quality at the receiver.

This paper is organized as follows. Section II describes our proposed method. Section III discusses the experimental results. Finally, section IV presents our conclusion.

II. PROPOSED RECEIVER-BASED VOIP QUALITY ENHANCING METHOD

A. *Structure of the Receiving Part in VoIP System*

The proposed structure of the receiving part of a mobile Internet phone is illustrated in Fig. 1. The receiving system employs combined signal reconstruction and playout control (SRPC) on the decoded signal frames.

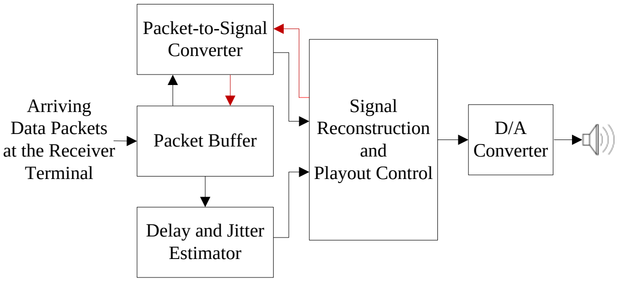

Packet-to-Signal Converter
Arriving Data Packets at the Receiver Terminal
Packet Buffer
Delay and Jitter Estimator
Signal Reconstruction and Playout Control
D/A Converter

Figure 1. Structure of the receiving part of a mobile Internet phone.

Arriving voice data packets from the sender over IP network is adequately placed in a packet buffer. To feed arriving packets to packet-to-signal converter at regular intervals, the receiving system needs to maintain a packet buffer. In response to the arriving packets in the packet buffer, the network jitter is adaptively estimated and used to assign each packet a controlled playout time in the SRPC module. The packet buffer holds incoming packets and then releases them for decoding at a regulated speed (i.e., every 10 ms), thereby reducing system delay. In this paper, the length of each packet is 20 ms, and the size of the packet buffer as a storage medium is 200 ms. Therefore, 1–10 packets are present in the packet buffer.

The decoded signal frames are entered into the SRPC module. The current used packet-to-signal converter is G.729.1 decoder.

In the SRPC, one of three processing modes (loss concealment, smoothing, or timing recovery) is performed for recovering lost packets and controlling the playout time on the decoded signal frames.

To recover lost packets, the SRPC module often makes a subsequent frame demand from the packet-to-signal converter, causing the packet-to-signal converter to make a packet demand from the packet buffer. The packet buffer then extracts a voice or music data packet and sends it to the packet-to-signal converter, which decodes it as a signal frame. The digital to analog converter (D/A) regularly converts the sampled signal frame from SRPC into an analog signal. Finally, the user hears the analog voice or music signal through a speaker.

B. Adaptive Playout Control and Signal Reconstruction

Fig. 2 presents an overall flow chart of three processing modes in the SRPC shown in Fig. 1.

After silence segments are discriminated between signal frames coming from the signal frame buffer, the SRPC performs one of three modes on the *i*-th signal frame (where $i = 1, 2\ldots I$ denotes the time index when each voice packet is generated at the sending host or when each voice packet is played out through the speaker) and *k*-th arriving packet (where $k = 1, 2\ldots K$ presents the number of packet at the receiving host) as follows:

1) Loss concealment mode

If the *i*-th signal frame (subsequent signal frame for playing out) is absent from the signal frame buffer, a packet is declared lost and the “loss concealment mode” is entered.

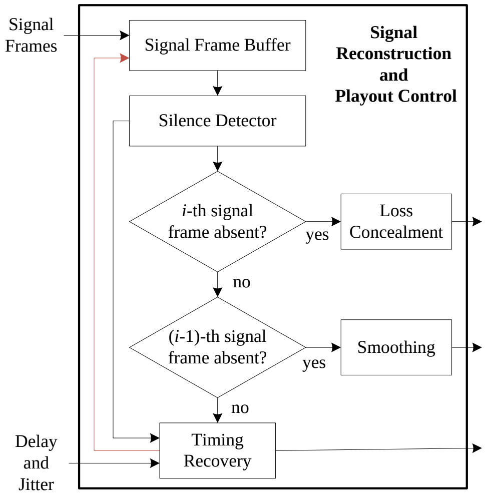


Figure 2. Flow chart of signal reconstruction and playout control.

Fig. 2 presents a flow chart of signal reconstruction and playout control. Our loss concealment algorithm is based on recursive LPAS using soft estimated pitch period to improve the G.722 Appendix IV PLC algorithm [5]. In the proposed recursive LPAS, the soft estimated pitch period is used to generate a smooth excitation for recovery of the lost packets, and effectively reduces tonal artificial frequencies that could be caused by repeating a small segment many times. If consecutive packets are lost, then the previous synthesized signal is recursively input to LP filter to generate the new smooth excitation signal. The reconstructed signal is synthesized by filtering the smooth excitation signal and gradually muted for the duration of the loss period.

2) Merging and smoothing mode

If the *i*-th signal frame is present in the signal frame buffer and the (*i*-1)-th signal frame was lost, discontinuity between the *i*-th signal frame and the (*i*-1)-th substituted signal frame occurs and the “smoothing mode” is entered. The smooth interpolation is obtained as follows: First, *N* samples from the *i*-th signal frame are obtained and input to the LPAS to generate the reference segment C. Second, the signal segment A most similar to the reference segment C is found in the samples of the (*i*-1)-th previous signal frame in the history buffer. Third, the smooth estimated signal frame D is generated using peak alignment overlap-add between the signal frame A and the signal frame C. Fourth, the (*i*-1)-th signal frame, the segment D, and the (*i*+1)-th signal frame are merged into the new segment O. The segment O is substituted into the *i*-th signal frame.

3) Timing recovery mode

If the *i*-th signal frame is present in the signal frame buffer and the (*i*-1)-th signal frame was not lost, the “timing recovery mode” is entered.

Fig. 3 depicts the algorithm flow chart of the decision of compression or normal playout process.

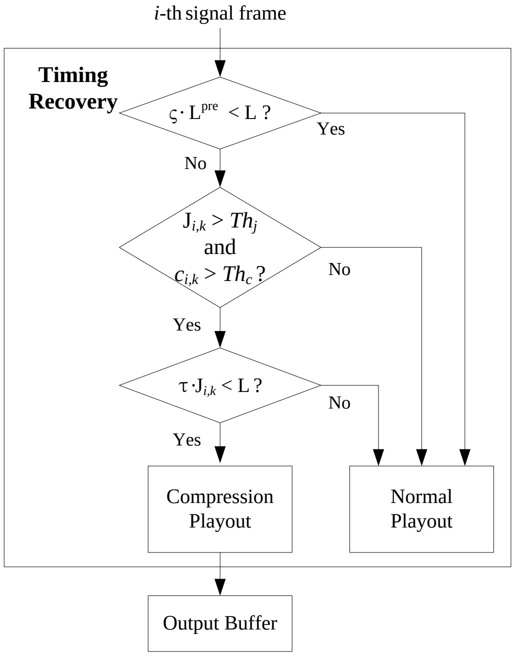


Figure 3. Flow chart of timing recovery.

The decision logic between normal playout or compression processes for the playout scheduling is performed using the estimated network jitter and the length of the remaining signal frames in the signal frame buffer. For this, the following subprocesses are handled:

- Let L and L^{pre} denote the total length of the current remaining signal frames and the previous remaining signal frames in the signal frame buffer, respectively. If L is larger than $ς\cdot L^{pre}$ ($1 \leq ς \leq 4$), the normal subprocess is initiated.
- If L is smaller than $ς\cdot L^{pre}$,, the estimated jitter $J_{i,k}$ is smaller than the jitter threshold Th_j ($10 \leq Th_j \leq 20$), and the jitter variance $c_{i,k}$ is smaller than the compression threshold Th_c ($3 \leq Th_c \leq 7$), then the normal subprocess is initiated.
- If the following three conditions are satisfied: $L < ς\cdot L^{pre}$; $J_{i,k} > Th_j$; and $c_{i,k} > Th_c$; and $\tau\cdot J_{i,k} \geq L$ using τ($2 \leq \tau \leq 5$), then the normal subprocess is initiated.
- If the following four conditions are satisfied: $L < ς\cdot L^{pre}$; $J_{i,k} > Th_j$; and $c_{i,k} > Th_c$; $\tau\cdot J_{i,k} < L$, then the compression subprocess is initiated.

C. Network Jitter Estimation

Our network jitter estimation incorporates spike detection and accurately predicts network delays including spike delays; thus, it is well-suited for timing recovery in playout algorithms in which playout delay is adjusted for each individual packet.

Fig. 4 depicts an algorithm flow chart of the proposed active jitter estimation.

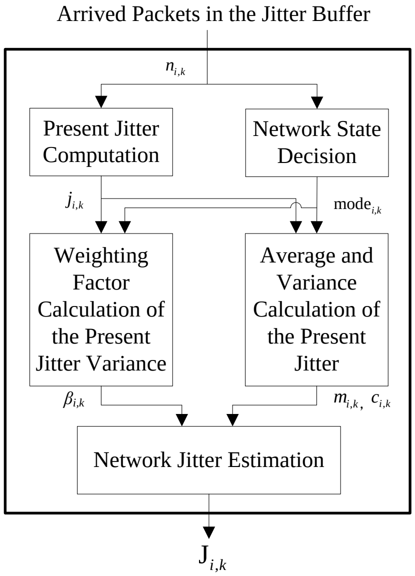

Arrived Packets in the Jitter Buffer
n_{i,k}
Present Jitter Computation
Network State Decision
j_{i,k}
mode_{i,k}
Weighting Factor Calculation of the Present Jitter Variance
Average and Variance Calculation of the Present Jitter
β_{i,k}
m_{i,k}, c_{i,k}
Network Jitter Estimation
J_{i,k}

Figure 4. Block diagram of the proposed network jitter estimation.

The active jitter estimation is composed of five modules: present jitter computation, network state decision, weighting factor calculation of the present jitter variance, average and variance calculation of the present jitter variance, and network jitter estimation.

The network jitter estimation procedure is as follows:

(Step1) Inter-arrival jitter results in a network delay change. Therefore, the inter-arrival jitter of the *k*-th arriving packet in the *i*-th signal interval is computed.

(Step2) Using a modified enhanced normalized least mean squares algorithm (E-NLMS) [1], network state is classified into one of two zones: "*spike*" or "*normal*." The modified E-NLMS algorithm is incorporated with delay spike detection using weighting factor $\beta_{i,k}$ of the network jitter variance. A delay spike is detected when the actual network delay exceeds the predicted delay value or the previous delay by a threshold. When the delays drop down to the level before the mode is in force, the normal mode is switched.

(Step3) The weighting factor $\beta_{i,k}$ of the inter-arrival jitter variance can be obtained as:

$$\text{If } (\text{mode}_{i,k} = normal)$$

$$\beta_{i,k} = \begin{cases} \beta_{i,k-1} - \alpha_c, \; where \; \beta_{i,opt} < \beta_{i,k-1} \\ \beta_{i,k-1} + \alpha_d, \quad otherwise \end{cases} \tag{1}$$

$$\text{else}$$

$$\beta_{i,k} = \beta_{i,k-1}$$

using α_c ($0 < \alpha_c < 1$), α_d ($0.5 < \alpha_d < 1.5$) and

$$\beta_{i,opt} = \frac{j_{i,k} - m_{i,k-1}}{c_{i,k-1}} \tag{2}$$

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-342-1

where $\beta_{i,opt}$ is a weighting factor of optimal network jitter variance to minimize the jitter error incurred by varying network conditions.

(Step4) After the adjustments of $\beta_{i,k}$, average $c_{i,k}$, and variance $m_{i,k}$ of the inter-arrival jitter are calculated according to the determined network situation, as shown in (3):

$$\begin{aligned}&\text{If } (\text{mode}_{i,k} = normal \text{ and } \text{mode}_{i,k-1} = spike)\\&\quad m_{i,k} = \alpha \cdot m_{i,tmp} + (1-\alpha)\cdot j_{i,k},\\&\quad c_{i,k} = \alpha \cdot c_{i,tmp} + (1-\alpha)\cdot \left|m_{i,k} - j_{i,k}\right| \qquad (3)\\&\text{else}\\&\quad m_{i,k} = \alpha \cdot m_{i,k-1} + (1-\alpha)\cdot j_{i,k}\\&\quad c_{i,k} = \alpha \cdot c_{i,k-1} + (1-\alpha)\cdot \left|m_{i,k} - j_{i,k}\right|\end{aligned}$$

where mode$_{i,k}$ and mode$_{i,k\text{-}1}$ represents the current and previous network state mode, respectively; α ($0 < \alpha < 1$) is a smoothing parameter; *tmp* is the temporal point when the spike is detected; and $m_{i,tmp}$ and $c_{i,tmp}$ are the mean and the variance of the jitter at the point at which the previous spike was detected, respectively.

(Step 5) The active network jitter of the *k*–th arriving packet in the *i*–th signal interval is estimated using the calculated means and variances of the inter-arrival jitters as:

$$\mathrm{J}_{i,k} = m_{i,k} + \beta_{i,k} \cdot c_{i,k} \qquad (4)$$

The estimate for the network delay $En_{i,k}$ is computed as

$$En_{i,k} = \alpha_n \cdot En_{i,k-1} + (1-\alpha_n)\cdot n_{i,k} \qquad (5)$$

where α_n ($0 < \alpha_n < 1$) is a weighting factor that controls the algorithm convergence rate and $n_{i,k}$ is the network delay that the *k*-th transmitted packet experiences.

The playout times are then adjusted as

$$p_{i+1,k} = En_{i,k} + R_{i,k} + \tau \cdot \mathrm{J}_{i,k} \qquad (6)$$

where $R_{i,k}$ is the timing recovery delay, and τ ($0 < \tau < 3$) controls the additional buffering delay and lateness loss ratio.

III. Experimental Results

A. Testbed Infrastructure and Measurements

In order to evaluate the proposed adaptive playout control and signal reconstruction, a test bed is set up [14]. These are connected to each other by two types of networks: ethernet (100 Mbps) and wireless local area network (WLAN) (300 Mbps). SIP signaling involves a SIP Proxy Server and two *clients*. VoIP application is developed by using Visual C++ and is installed in *clients*. *Clients* are mobile devices that have the following specifications: 800 MHz CPU, and 4 GB memory. SIP signaling messages are transferred through the audio data transport module and are sent from clients to the SIP Proxy Server, and then the SIP signaling messages redirected to clients accordingly. Clients send audio data in RTP packets. In the audio data transport module, Clients A and B are each connected to access points (access points 1 and 2) of ipTime N604M (Hubs 1 and 2). In the network traffic emulator module, a traffic generator is used in order to simulate WLAN connections with different traffic loads such as delay, jitter, and packet loss.

The speech samples are digitized at 16 kHz. Each trace lasts for about 5 min and consists of 15,000 packets, each of which consists of 20 ms of speech content. The music samples consists of a database of 100 songs from different genres such as pop, hip-hop, jazz, and classical and are digitized at 16 kHz.

To evaluate the quality degradation, objective voice quality testing is performed using perceptual evaluation of speech quality (PESQ), total buffering delay (TBE), and jitter estimation error (JER). PESQ is a recognized method for accurately testing the quality level that will be perceived by a VoIP network user and is described in the latest ITU-T recommendation P.862 Amendment 2 [15]. PESQ provides a score ranging from 1 to 5, where 1 is unacceptable and 5 is excellent. A typical range for VoIP is 3.5 to 4.2. TBE and JER are defined as follows:

$$\mathrm{TBE} = \sum_{k=1}^{\mathrm{K}} \left(p_{i,k} - a_{i,k}\right)/\,\mathrm{K}, \qquad (10)$$

$$\mathrm{JER} = \sum_{k=1}^{\mathrm{K}} \left(J_{i,k} - \mathrm{J}_{i,k-1}\right)/\,\mathrm{K} \qquad (11)$$

B. Comparison Results of Jitter Estimation based on Spike Detection

To evaluate the performance of the proposed jitter estimation based on spike detection, the experiment was performed on four network traces listed in Table 1.

TABLE I. Statistics of Network Traces

Trace	End-to-end network delay (ms)	STD of network delay (ms)	Maximum jitter (ms)	Network packet loss (%)
A	49.29	26.02	295	0
B	42.17	57.75	392	0
C	48.79	31.97	374	0

In Table I, average of the network delay, standard deviation (STD) of the network delay (which reflects the jitter characteristics for each condition), and maximum jitter (which is the difference between the maximum and minimum delay in the short trace) are depicted. Because we want to focus on the effect of jitter estimation based on spike detection in this section, four network delay traces with the extreme maximum jitter over 250 ms are chosen from the

Internet links of the testbed infrastructure. And the network traces do not carry any network packet loss.

The performance of the proposed jitter estimation method is compared with three methods. The three methods used have been modified from the contents of reference papers and implemented. Method 1 is based on the adaptive gap-based algorithm [7] incorporated with spike detection [3], while Method 2 is based on an adaptive NLMS playout algorithm with delay spike detection [1]. In Method 3, a timing recovery and loss substitution method [16] is combined with modeling the statistics of the interarrival times with the K-Erlang distribution [5]. Four methods commonly incorporate the packet loss concealment [10].

Table II shows the experimental results on jitter estimation error and late loss rate for each trace. PM denotes the proposed algorithm.

TABLE II. PERFORMANCE OF THE JITTER ESTIMATION

Trace	Method	Jitter estimation error(ms)	Late loss rate (%)
A	PM	34.4	1.06
	Method 1	68.6	2.37
	Method 2	51.8	1.73
	Method 3	60.5	2.02
B	PM	46.1	1.87
	Method 1	86.9	3.22
	Method 2	67.5	2.53
	Method 3	77.3	2.89
C	PM	36.9	1.46
	Method 1	74.6	2.55
	Method 2	55.8	2.11
	Method 3	66.2	2.43

As shown in Table III, the proposed jitter estimation based on spike detection achieves smaller jitter-estimation errors and late loss rate overall than Method 1, Method 2, and Method 3.

C. Comparison Results of Adaptive Signal Reconstruction and Playout Control

The six network delay traces that we collected from the Internet links for the performance evaluation are listed in Table III.

TABLE III. STATISTICS OF NETWORK TRACES

Trace	End-to-End Network Delay (ms)	STD of Network Delay (ms)	Maximum Jitter (ms)	Network Packet Loss (%)
1	25.38	7.46	48	1.93
2	24.82	8.30	48	3.99
3	34.80	13.37	152	3.97
4	47.17	17.88	195	1.99
5	79.98	29.62	363	1.96
6	78.22	31.22	371	3.97

STD, standard deviation

The performance of our proposed method is compared with three methods that have been modified from the contents of reference papers and then implemented. Method 1 is based on an adaptive normalized least mean square playout algorithm with delay spike detection [1] and packet loss concealment [16]. In Method 2, the time-scale modification and loss substitution method [5] are combined with modeling the statistics of the inter-arrival times with the K-Erlang distribution [5].

Table IV depicts the experimental results of the four methods. M1, M2, and PM denote Method 1, Method 2, and the proposed method, respectively.

TABLE IV. EXPERIMENTAL RESULTS OF THE FOUR METHODS

Trace	Method	Speech (Sampling Rate: 16 kHz)			Music (Sampling Rate: 16 kHz)		
		TBE (ms)	JER (ms)	PESQ Score	TBE (ms)	JER (ms)	PESQ Score
1	M1	42.74	18.46	2.857	43.47	17.54	2.825
	M2	35.23	23.57	3.635	36.27	24.29	3.156
	PM	**26.48**	**14.17**	**3.902**	**28.67**	**14.23**	**3.432**
2	M1	44.51	18.62	2.703	45.64	20.09	2.512
	M2	31.66	23.41	3.243	33.36	23.27	2.753
	PM	**25.54**	**14.13**	**3.682**	**27.93**	**13.83**	**3.285**
3	M1	43.28	29.54	2.534	40.78	30.21	2.072
	M2	45.39	29.36	2.877	46.48	29.37	2.532
	PM	**40.36**	**21.47**	**3.432**	**39.74**	**21.38**	**3.125**
4	M1	41.54	29.85	2.821	41.36	28.02	2.356
	M2	50.15	29.19	3.267	49.11	28.11	2.943
	PM	**43.32**	**21.93**	**3.753**	**44.54**	**21.48**	**3.557**
5	M1	50.56	77.82	2.902	46.31	75.31	1.747
	M2	85.55	78.24	3.082	87.26	78.04	2.675
	PM	**59.86**	**68.18**	**3.414**	**66.19**	**68.84**	**3.237**
6	M1	51.81	78.51	2.679	45.11	78.55	1.747
	M2	85.47	78.18	2.605	78.78	78.25	2.248
	PM	**52.21**	**69.51**	**3.126**	**62.36**	**69.09**	**2.846**

TBE, total buffering delay; JER, jitter estimation error; PESQ, perceptual evaluation of speech quality; M1, Method 1; M2, Method 2; PM, proposed method

Table IV shows that our proposed method (PM) outperforms the reference methods M1, and M2 in medium jitter, high jitter, 2 % packet loss rate, and 4 % packet loss rate. The highest PESQ scores were achieved using the PM in traces 1. The PESQ scores of speech samples were higher than those of music samples.

In particular, as the jitter levels or packet loss rates increase, the PESQ scores decrease. The performance difference becomes more significant, showing the clear advantage of the PM. The PM is well-suited for operating with low buffering delay against dynamic changes in network conditions, and handling various loss patterns.

IV. CONCLUSIONS

In this paper, we proposed and evaluated an adaptive signal reconstruction and playout control for enhancing VoIP speech and music quality. The proposed fully receiver-based enhancing algorithm enables users to deliver high-quality voice or music using the combined signal reconstruction and playout control. Experimental results confirm that the proposed method achieves higher PESQ values than the other methods and is suitable for use in any practical mobile VoIP system.

ISBN: 978-1-61208-342-1

In the future, we will apply the proposed method to advanced teleconferencing applications running on Smart TV.

ACKNOWLEDGMENT

This research was supported by the MSIP(Ministry of Science, ICT & Future Planning), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency), NIPA-2013-H0301-13-4005

REFERENCES

[1] A. Shallwani, “An adaptive playout algorithm with delay spike detection for real-time VoIP,” Electrical and Computer Engineering, IEEE CCECE 2003, Canadian Conference on, Montreal, Quebec, Canada, vol. 2, May, 2003, pp. 997-1000.

[2] B. Sat and B. W. Wah, "Analyzing voice quality in popular VoIP applications," IEEE Multimedia, vol.16, issue 1, January, 2009, pp.46-59.

[3] Y. Liang, N. Farber, and B. Girod, “Adaptive playout scheduling and loss concealment for voice communication over IP networks,” IEEE Transactions on Multimedia, vol. 5, no. 4, April, 2001, pp. 532-543.

[4] V. P. Bhute and U. N. Sharawankar, “Speech packet concealment techniques based on time-scale modification for VoIP,” International Conference on Computer Science and Information Technology, Singapore, Singapore, August, 2008, pp. 825-828.

[5] H. Li, G. Zhang, and W. Kleijn, “Adaptive playout scheduling for VoIP using the K-Erlang distribution,” The 2010 European Signal Processing Conference, Algborg, Denmark, August, 2010, pp. 1494-1498.

[6] H. Sanneck, A. Stenger, K. B. Younes, and B. Girod, “A new technique for audio packet loss concealment,” Global Telecommunications Conference, London, UK, November 1996, pp. 48–52.

[7] H. Sanneck, A. Stenger, K. B. Younes, and B. Girod, “A new technique for audio packet loss concealment,” Global Telecommunications Conference, London, UK, November 1996, pp. 48–52,

[8] J.-H. Chen, “Packet loss concealment for predictive speech coding based on extrapolation of speech waveform,” ACSSC 2007 Conference Record of the Forty-First Asilomar Conference on Signal, Systems and Computers, California, USA, November 2007, pp. 2088-2092.

[9] J. Pinto and K. J. Christensen, “An algorithm for playout of packet voice based on adaptive adjustment of signalspurt silence periods,” 24th Conference on Local Computer Networks, Lowell, USA, vol. 5, October 1999, pp. 224-231.

[10] S. Chi and B. F. Womack, “QoS-based optimal adaptive playout buffer scheduling using the packet arrival distribution,” IEEE MILCOM, California, USA, October 2009, pp. 15-19.

[11] J. Aragao Jr. and G. Barreto, “Novel approaches for online playout delay prediction in VoIP applications using time series models,” Computers & Electrical Engineering, vol. 36, issue 3, May 2010, pp. 536-544

[12] D. Florencio and L.-W. He, “Enhanced adaptive playout scheduling and loss concealment techniques for Voice over IP networks,” 2011 IEEE International Symposium on Circuits and Systems, Rio de Janeiro, Brazil, May 2011, pp. 129-132.

[13] S. L. Ng, S. Hoh, and D. Singh, “Effectiveness of adaptive codec switching VoIP application over heterogeneous networks,” 2nd International Conference on Mobile Technology, Applications and Systems, Guangzhou, China, November 2005, pp. 7-13.

[14] ITU-T Rec. P.862, “Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality qssessment of narrowband telephone networks and speech codecs,” International Telecommunication Union, Geneva, Switzerland, February 2001.

[15] N. Aoki, “A VoIP packet loss concealment technique taking account of pitch variation in pitch waveform replication,” Electronics and Communications in Japan, vol. 89, no. 3, March 2006, pp. 1-9

[16] S. V. Andrsen, W. B. Kleijn, and P. Sorqvist, “Method and arrangement in a communication system,” U.S. Patent 7 321 851, 2008.

An Easy and Efficient Grammar Generator for Understanding Spoken Languages

A Novel approach to develop a Spoken Language Understanding Grammar for Inflective Languages

Salvatore Michele Biondi, Vincenzo Catania, Ylenia Cilano, Raffaele Di Natale, Antonio Rosario Intilisano
Dipartimento di Ingegneria Elettrica Elettronica e Informatica
University of Catania
Catania, Italy
salvo.biondi@dieei.unict.it, vincenzo.catania@dieei.unict.it, ylenia.cilano@dieei.unict.it, raffaele.dinatale@dieei.unict.it,
aintilis@dieei.unict.it

***Abstract*— In a Spoken Dialog System, the Spoken Language Understanding component is able to recognize words that were previously included in its grammar. The development of a grammar is a very time-consuming and error-prone process, especially for the inflectional languages, because the developer must manually include all possible inflected forms of a word. As a consequence, the grammar definition files are long and hard to manage. This paper describes a solution for creating a semi-automatic Spoken Language grammar using a morphological generator.**

Keywords-Spoken Language Understanding; Natural Language Understanding; Spoken Dialog System; Grammar Definition.

I. INTRODUCTION

Grammar development in a Spoken Dialog System (SDS) is the process of specifying the words and patterns of words that a speech recognizer should be able to process. The system is able to recognize a user's utterance of "hello" only if that word is included in the grammar. Manual development of domain-specific grammar is time-consuming, error-prone and requires a significant amount of expertise. It is difficult to write a rule–set that has a good coverage of real data without making it intractable [1]. Writing domain-specific grammars is a major obstacle to a typical application developer. This specialization often does not cover any unspecified data and it often results in ambiguities [2].

With this purpose, we suggest that a semi-automatic method for generating grammar contents for inflectional languages is necessary. In order to circumvent the complexity associated with conventional methods for automatic grammar inference for Spoken language, we pursue a different avenue. More precisely, this paper focuses on the simplification of the writing of a grammar, by means of a method that automatically generates the inflected forms of its terms.

This is accomplished by introducing an intermediate grammar that helps generating a simpler and more compact grammar. The development process allows to obtain large amounts of grammar contents starting from a few rows of the intermediate grammar.

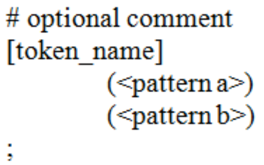
In order to test the validity of our solution, a specified grammar editor has been developed. It permits to automatically convert the new grammar format, developed in this work, in the Phoenix grammar [3]. Phoenix represent the Spoken Language Understanding (SLU) module of the Olympus Framework [4].

This paper is organized as follows: the Phoenix grammar format is described in Section II. The proposed grammar format in presented in Section III. Section IV introduces a Grammar generator based on a Morphological Generator for the Italian language and Section V shows an example. Finally, in Section VI, we draw conclusions.

II. PHOENIX GRAMMAR

The Phoenix parser represents the state of the art of the development of robust Spoken Language interfaces for spoken language applications. Spontaneous speech is often ill-formed and could cause recognition errors. For such a reason, the proposed parser is designed to enable robust parsing and is able to manage this kind of input. The Phoenix parser uses a specific grammar file (.gra) containing context free rules that specify the word patterns corresponding to the token.

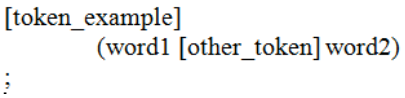
The pattern is a combination of words [5] that can be recognizable by the Phoenix parser. The syntax of a token is shown in Fig. 1:



```
# optional comment
[token_name]
        (<pattern a>)
        (<pattern b>)
;
```

Figure 1. Syntax of a token.

A token can also contain other tokens, as Fig. 2, for example:



```
[token_example]
        (word1 [other_token] word2)
;
```

Figure 2. Syntax token example.

ISBN: 978-1-61208-342-1

This format allows for recognition of several sentences with the combination of different slots and words; furthermore, each token can be reused in many tokens.

In the inflectional languages [6], as in the case of Italian or Romance languages in general, words can occur in several forms, verbs can change their form depending on conjugations and nouns and adjectives depending on declensions. Moreover, suffixes or prefixes can be applied to them.

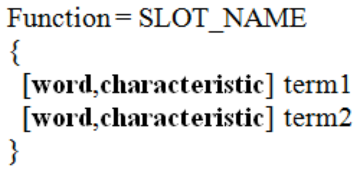
Thus, inflected forms add complexity to the Phoenix grammar, since they generate multiple different rules with similar patterns. That causes an increase of development time. Moreover, the developer might not include some inflected forms, thus causing the grammar to be incomplete.

III. A NEW GRAMMAR SCHEMA

The development of a new domain application needs a new Context Free Grammar (CFG) that is able to define the concepts and their relations of such domain.

Alternative approaches learn structures from a set of corpora. However, this process appears too expensive and potentially not exhaustive [7]. Our approach consists of creating a new intermediate grammar that focuses on the meaning of a grammar token rather than on its content.

In such a way, it is no longer necessary to write the word pattern of the token, but only the "keyword name" like [**word**,**characteristic**]. The new schema generates a grammar file containing a token and its generated word patterns and it can be reused and edited like a standard Phoenix grammar. The new format description is shown in Fig. 3:



```
Function = SLOT_NAME
{
 [word,characteristic] term1
 [word,characteristic] term2
}
```

Figure 3. New grammar format description.

The grammar slots [5] are defined by the "Function" keyword that defines the slot name (Function = SLOT_NAME). This way, a token is defined as a couple "[word, characteristic]" and is used by the editor to generate the appropriate **word** patterns according to the **characteristic**.

The couple [word,characteristic] is defined as below:

1) If "word" is a verb, "characteristic" can be replaced with:

a) "Presente" if the Italian language present form is desired;
b) "Passato" if the Italian language past forms are desired;
c) "Futuro" if the Italian language future forms are desired.

2) If "word" is a noun or an adjective, "characteristic" can be replaced with:

a) "Singolare" if the Italian language singular forms are desired;
b) "Plurale" if the Italian language plural forms are desired;

All new forms specified by the characteristic are generated by a Morphological Generator [8].

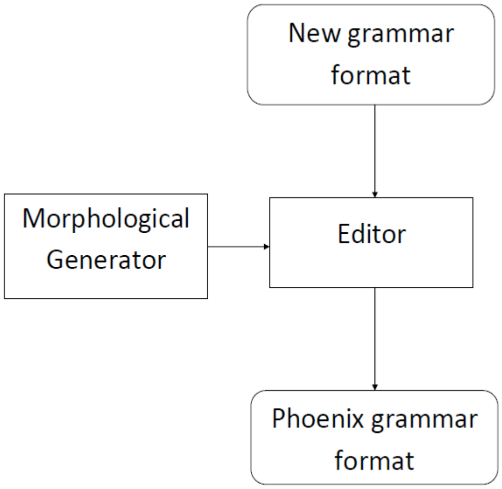


Figure 4. Grammar generation.

Our editor generates a standard Phoenix grammar from the new intermediate grammar, performing the following actions:

- Create a token named SLOT_NAME in which new tokens and terms are included;
- Create a token for each new defined token, in which terms generated by the Morphological Generator are included. The entire process is shown in Fig. 4.

IV. GRAMMAR GENERATOR

In our test, we used Italian as inflectional language, but different Romance languages can be used. Each inflected form of a verb gives information about mood, tense, number and person. There are also some verbs, nouns and adjectives that are inflected in an irregular manner. In Italian, nouns and adjectives can be altered by adding particular suffixes.

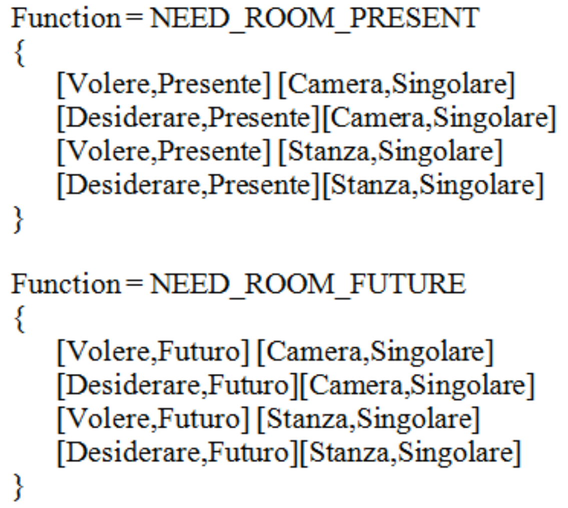
These alterations modify a word's meaning in terms of quantity or quality. This increases the effort in developing an efficient Spoken Language Understanding grammar for a SDS.

In our solution, a Morphological Generator generates all inflected forms of a word for Italian language. Its aim is to help the programmer to generate a complete SLU grammar for a SDS in an easy way.

V. EXPERIMENTAL RESULTS

An example is reported to show the advantages obtained by this approach. It shows a grammar developed for a room reservation application. In a typical interaction, the user can express the same concept using a specific word, but in different tenses.

For example, “I want a room” in Italian can be expressed like “Voglio una camera” but also “Vorrei una camera” (I'd like to have a single room.) or “Vorrei una cameretta” (I'd like to have a small room.). Fig. 5 shows an example of grammar:

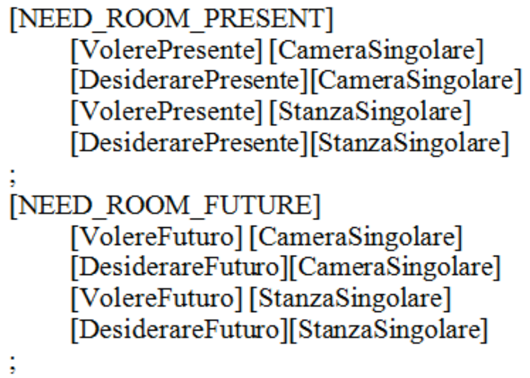


```
Function = NEED_ROOM_PRESENT
{
    [Volere,Presente] [Camera,Singolare]
    [Desiderare,Presente][Camera,Singolare]
    [Volere,Presente] [Stanza,Singolare]
    [Desiderare,Presente][Stanza,Singolare]
}

Function = NEED_ROOM_FUTURE
{
    [Volere,Futuro] [Camera,Singolare]
    [Desiderare,Futuro][Camera,Singolare]
    [Volere,Futuro] [Stanza,Singolare]
    [Desiderare,Futuro][Stanza,Singolare]
}
```

Figure 5. New grammar format example.

The new grammar consinsts of two parts. The first one, shown in Fig. 6, represents the definition of a grammar slot:



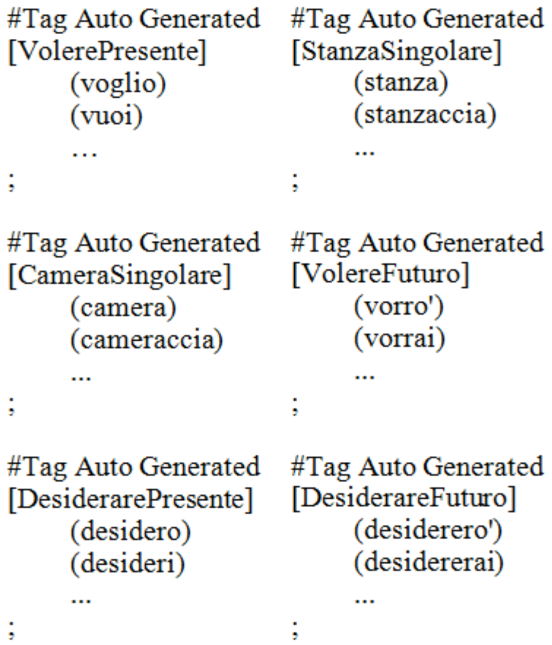
```
[NEED_ROOM_PRESENT]
        [VolerePresente] [CameraSingolare]
        [DesiderarePresente][CameraSingolare]
        [VolerePresente] [StanzaSingolare]
        [DesiderarePresente][StanzaSingolare]
;
[NEED_ROOM_FUTURE]
        [VolereFuturo] [CameraSingolare]
        [DesiderareFuturo][CameraSingolare]
        [VolereFuturo] [StanzaSingolare]
        [DesiderareFuturo][StanzaSingolare]
;
```

Figure 6. Phoenix grammar generated.

The second part, shown in Fig. 7, defines each token including their word patterns. A more detailed explanation along with the source code (output.gra file) is given in [9].

The initial grammar, consisting of 21 rows, generates a 140-row-long Phoenix grammar that allows the SLU module to recognize a large set of utterances.

This way, the developer focuses his attention on the meaning of an intermediate-grammar token and not on its content.



```
#Tag Auto Generated        #Tag Auto Generated
[VolerePresente]           [StanzaSingolare]
        (voglio)                   (stanza)
        (vuoi)                     (stanzaccia)
        …                          ...
;                          ;

#Tag Auto Generated        #Tag Auto Generated
[CameraSingolare]          [VolereFuturo]
        (camera)                   (vorro')
        (cameraccia)               (vorrai)
        ...                        ...
;                          ;

#Tag Auto Generated        #Tag Auto Generated
[DesiderarePresente]       [DesiderareFuturo]
        (desidero)                 (desiderero')
        (desideri)                 (desidererai)
        ...                        ...
;                          ;
```

Figure 7. Token definition generated.

Furthermore, the developer does not need to write all possible forms (mood, tense, person, etc.), some of which could be difficult to predict. The advantage of the generated grammar is the ability to easily simulate and predict the large variety of interactions that can occur.

VI. CONCLUSION AND FUTURE WORK

This paper proposed a solution to simplify and reduce the amount of writing of the SDS grammar of inflectional language. This method reduces the effort to produce a grammar for a SDS. The SDS used for our tests is the Olympus framework.

An editor has been developed for the translation of the new simple grammar format in the Phoenix grammar format. The editor uses a Morphological Generator to obtain all possible inflected words that are used to create grammar tokens.

The proposed solution will be integrated in our major project called Olympus P2P [10], which is concerned with the upgrading and updating of an SDS grammar by means a Peer to Peer Network to share new grammar tokens.

ACKNOWLEDGMENT

The authors were supported by the Sicilian Region grant PROGETTO POR 4.1.1.1: "Rammar Sistema Cibernetico programmabile d’interfacce a interazione verbale".

REFERENCES

[1] H. M. Meng and K-C. Siu, “Semtiautomatic Acquition of Semantic Structures for Understanding Domain-Specific Natural Language Queries”, IEEE Tran. Knowledge & Data Eng., pp. 172–181, vol. 14(1), 2002.

[2] Y. Wang and A. Acero, "Grammar learning for spoken language understanding," *Automatic Speech Recognition and Understanding, 2001. ASRU '01. IEEE Workshop on*, pp.292-295, 2001.

[3] W. Ward, "Understanding spontaneous speech: the Phoenix system," *Acoustics, Speech, and Signal Processing, 1991. ICASSP-91, 1991 International Conference on*, 14-17 Apr 1991, pp.365-367 vol.1.

[4] D. Bohus, A. Raux, T. K. Harris, M. Eskenazi and A. I. Rudnicky, *Proceedings of the Workshop on Bridging the Gap: Academic and Industrial Research in Dialog Technologies*, NAACL-HLT-Dialog '07 (Association for Computational Linguistics, Stroudsburg, PA, USA, 2007), pp. 32–39.

[5] Phoenix Parser User Manual, http://www.ontolinux.com/community/phoenix/Phoenix_Manual.pdf (last visited: 22 November 2013).

[6] M. Haspelmath and A. D. Sims, Understanding Morphology 2nd edition. London: Hodder Education, 2010.

[7] S. Knight, G. Gorrell, M. Rayner, D. Milward, R. Koeling and I. Lewin, "Comparing grammar-based and robust approaches to speech understanding: a case study", EUROSPEECH 2001 Scandinavia, 7th European Conference on Speech Communication and Technology, 2nd INTERSPEECH Event, Aalborg, Denmark, September 3-7, 2001, pp. 1779-1782.

[8] V. Catania, Y. Cilano, R. Di Natale, V. Mirabella and D. Panno, "A morphological engine for Italian language", ICIEET 2013: 2nd International Conference on Internet, E-Learning & Education Technologies, 2013, pp. 36-43, vol. 12(1).

[9] Source code, http://opensource.diit.unict.it/vctsds/GrammarEditor.zip (last visited: 22 November 2013).

[10] V. Catania, R. Di Natale, A. Longo and A. Intilisano, "A distributed Multi-Session Dialog Manager with a Dynamic Grammar Parser", 2nd International Conference on Human Computer Interaction & Learning Technologies, 2013, pp. 1-9, vol. 8(2).

RPKOM-GEN

A System for Testing Speech Recognition in Adverse Acoustic Conditions Using Speech Synthesis

Marián Trnka, Milan Rusko, Sakhia Darjaa, Róbert Sabo, Juraj Pálfy, Štefan Beňuš, Marian Ritomský, Martin Dravecký
Institute of Informatics, Slovak Academy of Sciences, Bratislava, Slovakia
{marian.trnka, milan.rusko, utrrsach, robert.sabo, juraj.palfy, stefan.benus, marian.ritomsky, martin.dravecky}@savba.sk

***Abstract*—Training and testing of current state-of-the-art speech recognition systems require huge speech databases whose creation is time-consuming and expensive. This paper presents a novel approach for testing speech recognition in adverse acoustic conditions that uses speech synthesis, which facilitates optimizing and adjusting speech recognition to various environmental conditions. RPKOM-GEN is a complex system of multiple synthesizers that generates synthetic speech and testing signals with well defined characteristics. It might be used to produce public announcements, sets of utterances for spoken dialogue systems or other speech excerpts. The acoustic parameters of synthetic voices, such as speech rate, pitch, intensity, and others, can be pre-defined from a broad range of options. By using this novel technique, the system can also vary vocal effort imitating thus the Lombard effect and so-called long-distance speech. It is also possible to model the characteristics of the transmission channel since the system includes noise generators and digital effects such as the setting of environmental noise or reverberation levels. The paper presents the system architecture, describes graphical user interface and a rich array of usage possibilities, and discusses the results of pilot experiments testing the effect of added noise on speech recognition accuracy.**

Keywords-speech recognition; adverse conditions; noise; speech synthesis.

I. Introduction

One of the most demanding tasks in research and development of automatic speech recognition applications employed in situations with the transmission channel difficulties are the preparation, elicitation, annotation, and processing of speech databases. In an ideal case, high-quality databases should include speech from multiple and varied speakers recorded in real communicative situations under various physical (i.e., acoustic) conditions of the environment and the channel. The recording should be statistically representative in the sense that they should cover many combinations of factors under which a speech recognition system might be deployed. The variability of the factors, and consequently their combinations, is, however, enormous and creating a database that would cover all of them is, in effect, impossible.

One of the possible ways of approaching this problem is to substitute the recordings of real speech with synthesized acoustic signals, which allows imitating of various factors on the transmission channel with the help of synthesized signals and by introducing various effects through digitally processing the signal. This approach enables the creation of huge number of speech signals that can be used not only to verify understandability of a particular speech synthesis system in noise conditions but also to test the efficiency of a speech recognition system under various levels and types of noise present in the environment compounded with various characteristics of the transmission channel. Besides testing synthetic voices, the approach that exploits Text To Speech (TTS) synthesis also provides an option to set the characteristics of a particular synthetic voice, and imitate thus the changes speakers make in adverse acoustic conditions.

The paper contributes to the Activity 3.3 “Automatic speech recognition in adverse environments“, which is included in the EU-funded project “Technology research for the management of business processes in heterogeneous distributed systems in real time with the support of multimodal communication“ – RPKOM (acronym RPKOM is a short-hand for the project name in Slovak). The goal of the activity is to conduct applied research in automatic speech recognition for adverse acoustic environments and propose algorithms and architecture for systems capable of 1) generating announcements of public information systems and utterances of spoken dialogue systems that are reliably understandable by humans, 2) recognizing spoken instructions in noisy environments, 3) synthesizing speech that is optimized to achieve high understandability in highly noisy environments, and 4) being included in a speech recognition system for Slovak that is robust in dealing with acoustic environmental noise and varied characteristics of the transmission channel.

Many authors have been trying to find methods to improve intelligibility of speech synthesis in noisy and reverberant environments [1][2][3]. Noisy and reverberant environments represent an issue also for speech recognition [4][5][6]. We, therefore, decided to develop a tool that will be able to generate synthesized speech signals, mix them with various noises and apply reverberation. This tool will be used for experiments with speech synthesis and speech recognition in simulated adverse acoustic conditions.

In this paper, we start by presenting a simplified model of speech communication that informed the design of RPKOM-GEN in Section II. The system architecture is described in

Section III. Section IV sketches the design of graphical user interface. Pilot experiments testing the effect of adding various types and levels of noise on speech recognition accuracy are discussed in Section V. Section VI concludes the paper.

II. Simplified Model of Speech Communication

A detailed description of spoken communication using a complex model with a range of influencing factors is beyond the scope of this paper; see, for example, [7]. For our purposes, a simplified model is sufficient for achieving functional solutions that can be implemented and deployed in real life. Such a simplified model is depicted in Fig. 1 and we will briefly discuss its components. We limit the discussion to the first two components – the speaker, the channel, and the factors influencing them – since they are most closely linked to the core of the proposed approach.

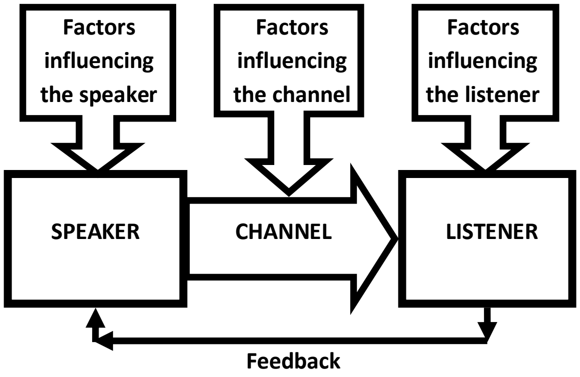

Factors influencing the speaker
Factors influencing the channel
Factors influencing the listener
SPEAKER
CHANNEL
LISTENER
Feedback

Figure 1. A schematic illustration of a simplified model of uni-directional transmission of information through the speech channel.

A. *Speaker (Message sender)*

The speaker sends the message through the acoustic signal produced by the articulatory speech production process. The characteristics of this signal are affected by multiple factors, some of which depend primarily on the speaker:

- Linguistic factors, such as the semantic content of the message, features of the lexicon, grammatical structuring, style, and others
- Paralinguistic factors, such as disfluencies, emotional states, and others
- Extra-linguistics, factors such as age, gender, speaker’s health condition, and others.

B. *Factors influencing the speaker*

The speech signal is the primary carrier of information. Some external factors might induce changes in speaker’s abilities, physical and mental conditions, or decision making processes. These factors then influence the final characteristics of the speech signal. In the RPKOM project, we focus mostly on the influence of external factors that increase speaker’s stress.

C. *Channel*

In this paper, we consider the channel to represent the entire transmission process that the speech signal must undergo from speech generation by the speaker to speech decoding by the listener. We are thus concerned with the propagation of sound through air in some acoustic environment. Furthermore, we also include here the process of tracking the sound with a microphone, digitizing the analogue signal, coding, transmission of some telecommunication channel, such as internet cable, decoding, digital to analog conversion, and playback through speakers or headphones. Note that sound propagation through air is also involved when sound travels from the loudspeakers to the listener.

D. *Factors influencing the channel*

When the speech signal travels through the acoustic environment, is can be affected by the properties of barriers against which it bounces, or the presence and characteristics of background noise. Within the telecommunication channel, the quality of the signal is degraded by the noise of the channel itself and by the processes of digitizing and coding. There might also be signal distortions specific to a particular telecommunication transmission channel, such as delays or missing packets.

Our system covers primarily three types of acoustic signal degradation:

- Adverse influence of the acoustic environment such as acoustic bounces, reverberations, echo, and others
- Noise and non-speech sounds in the background such as pink, white, bubble, or cocktail-party noise
- Speech of other speakers

Other specific aspects of signal transmission such as microphone overload, specific factors of customer transmission channel, packet drop-outs, and others are left for future research.

III. Rpkom-Gen Architecture

The architecture of the system is sketched in Fig. 2. The core of the RPKOM-GEN system is the *Signal Processing Unit* that has a functionality of mixing the speech signal with noise of various types while controlling for the signal-to-noise ratio. The database of noises (*Noise DB*) contains noise samples. The signal might be further distorted by adding the effects simulating adverse acoustic conditions such as *Delay*, *Reverberation*, *Echo*, and others.

The input of the signal processing unit is the speech signal that comes either from recorded human speech (*Speech Recordings*) or from synthesized speech produced by the *Speech Synthesis Unit*. This unit includes two types of synthesizers. The first one, *Unit TTS*, is based on corpus speech synthesis which selects and consequently joins the most suitable units of speech found in the speech corpus *Unit DB* [8][9]. Te second type of speech synthesis, *HMM TTS*, is based on Hidden Markov Models [10][11]. These statistic

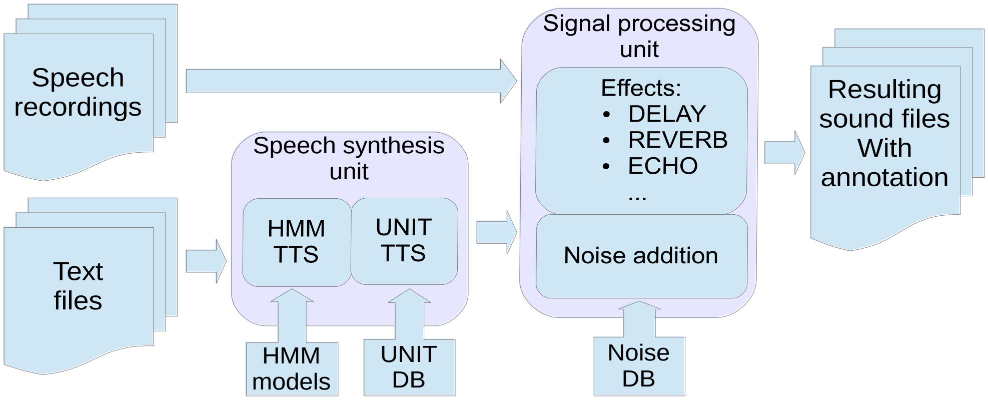


Figure 2. RPKOM-GEN system architecture.

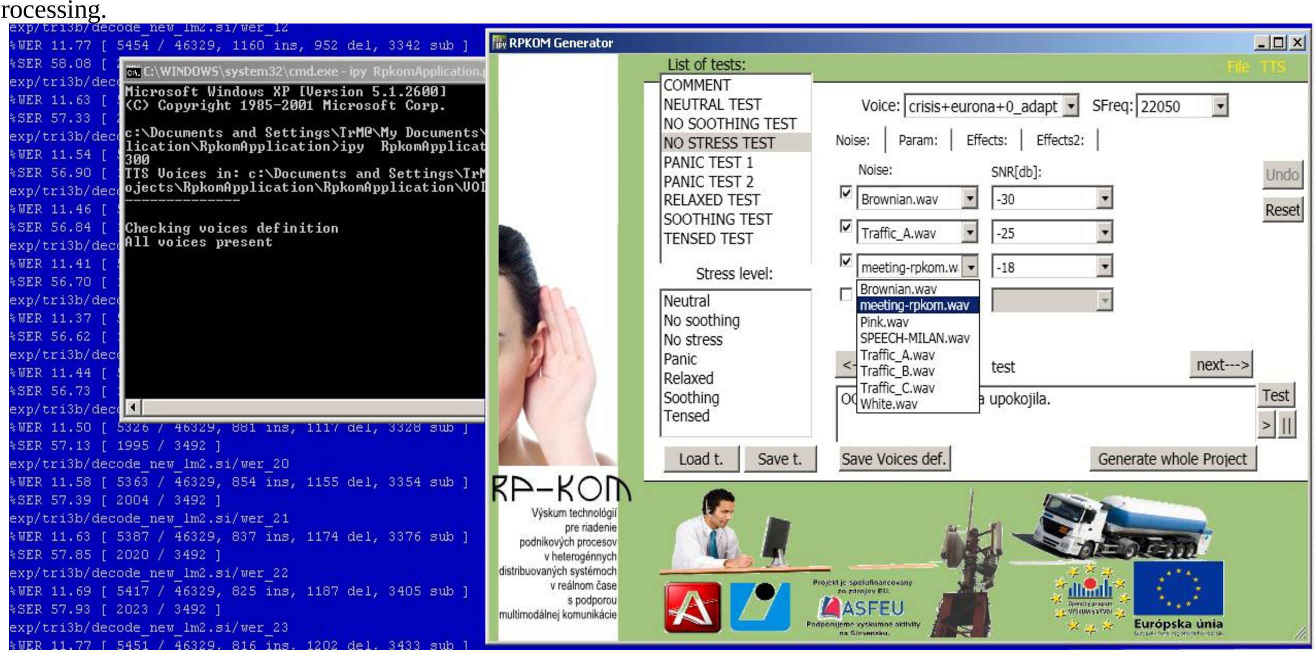
parametric synthesizers produce the models of acoustic parameters (*HMM Models*). One of the differences between the two synthesizer types, that is relevant for this paper, is the control of speech effort, which is extremely important for synthesizing shouted speech and simulating the speech with the Lombard effect [12]. The Unit TTS offers the control of the speaker voice, the prosody model, primarily covering the rhythm and intonation of speech, mean fundamental frequency (F0), mean speech rate, and F0 range. The speech of HMM TTS voices, in addition to the above control parameters, also allows for manipulating speech effort exploiting our novel method for expressive speech synthesis [13]. Finally, the input for the text-to-speech synthesis is the set of pre-defined instructions and other texts contained in the database *Text Files*.

The system output is represented by *Resulting Sound Files* that contain detailed *Annotations* describing the content and the settings of all parameters applied during signal processing.

IV. USER INTERFACE

An example of graphical user interface design is shown in Fig. 3. The user interface is implemented in the Iron Python scripting environment. The user first selects whether real or synthesized speech should be used as the input.

In the latter, the user creates a *project* that collects all texts to be synthesized. The user then pairs each text with its own name of the testing voice. When prompted, the system reads in the project, lists all the testing voices (*List of tests*), and offers a list of available pre-defined synthetic voices (*Voice*). For each testing voice, i.e., for each *test*, the user selects from the available synthetic voices the one that is closest to his/her requirements and subsequently adjusts the individual parameters for signal processing in the Parameter panel (*Param*).



```
exp/tri3b/decode_new_lm2.si/wer_12
%WER 11.77 [ 5454 / 46329, 1160 ins, 952 del, 3342 sub ]
%SER 58.08 [
exp/tri3b/dec
%WER 11.63 [
%SER 57.33 [
exp/tri3b/dec
%WER 11.54 [
%SER 56.90 [
exp/tri3b/dec
%WER 11.46 [
%SER 56.84 [
exp/tri3b/dec
%WER 11.41 [
%SER 56.70 [
exp/tri3b/dec
%WER 11.37 [
%SER 56.62 [
exp/tri3b/dec
%WER 11.44 [
%SER 56.73 [
exp/tri3b/dec
%WER 11.50 [ 5326 / 46329, 881 ins, 1117 del, 3328 sub ]
%SER 57.13 [ 1995 / 3492 ]
exp/tri3b/decode_new_lm2.si/wer_20
%WER 11.58 [ 5363 / 46329, 854 ins, 1155 del, 3354 sub ]
%SER 57.39 [ 2004 / 3492 ]
exp/tri3b/decode_new_lm2.si/wer_21
%WER 11.63 [ 5387 / 46329, 837 ins, 1174 del, 3376 sub ]
%SER 57.85 [ 2020 / 3492 ]
exp/tri3b/decode_new_lm2.si/wer_22
%WER 11.69 [ 5417 / 46329, 825 ins, 1187 del, 3405 sub ]
%SER 57.93 [ 2023 / 3492 ]
exp/tri3b/decode_new_lm2.si/wer_23
%WER 11.77 [ 5451 / 46329, 816 ins, 1202 del, 3433 sub ]
```

C:\WINDOWS\system32\cmd.exe - ipy RpkomApplication.

```
Microsoft Windows XP [Version 5.1.2600]
(C) Copyright 1985-2001 Microsoft Corp.

c:\Documents and Settings\TrM@\My Documents\
lication\RpkomApplication>ipy  RpkomApplicat
300
TTS Voices in: c:\Documents and Settings\TrM
ojects\RpkomApplication\RpkomApplication\VOI
---------------

Checking voices definition
All voices present
```

RPKOM Generator

List of tests:
COMMENT
NEUTRAL TEST
NO SOOTHING TEST
NO STRESS TEST
PANIC TEST 1
PANIC TEST 2
RELAXED TEST
SOOTHING TEST
TENSED TEST

Stress level:
Neutral
No soothing
No stress
Panic
Relaxed
Soothing
Tensed

Voice: crisis+eurona+0_adapt
SFreq: 22050

Noise:	SNR[db]:
☑ Brownian.wav	-30
☑ Traffic_A.wav	-25
☑ meeting-rpkom.w	-18
☐	

Brownian.wav
meeting-rpkom.wav
Pink.wav
SPEECH-MILAN.wav
Traffic_A.wav
Traffic_B.wav
Traffic_C.wav
White.wav

test

OC a upokojila.

File
TTS
Noise:
Param:
Effects:
Effects2:
Undo
Reset
<-
next--->
Test
>
||
Load t.
Save t.
Save Voices def.
Generate whole Project
RP-KOM
Výskum technológií pre riadenie podnikových procesov v heterogénnych distribuovaných systémoch v reálnom čase s podporou multimodálnej komunikácie
Projekt je spolufinancovaný zo zdrojov EÚ.
ASFEU
Podporujeme výskumné aktivity na Slovensku.
Európska únia

Figure 3. RPKOM-GEN system interface

Panel Noise offers a selection of noise types and the user might also control the resulting signal-to-noise ratio. Finally, the user might choose a type of a digital effect and the setting of its parameters in the Effects panel.

V. EXPERIMENTS

In this section, we present the results of pilot experiments testing the effect of added noise on speech recognition accuracy.

A. *Methodology: adding noise*

Depending on the setting of the parameters described in previous sections, mixing of noise components in the input signal is illustrated in Fig. 4. In the first step, the root mean square (RMS_S) value of the original acoustic signal is calculated using only those intervals that the *Voice Activity Detector* identifies as speech. The same method is applied for calculating RMS_N of the selected noise signal. Based on the ratio of the two RMS values and the selection of the required Sound-to-Noise Ratio (SNR), we calculate the coefficients for weighting the original and the noise signals. All processing of the signals is done on 32-bit-integer numeric fields to prevent overflow. Finally, the resulting signal is normalized to 16-bit.

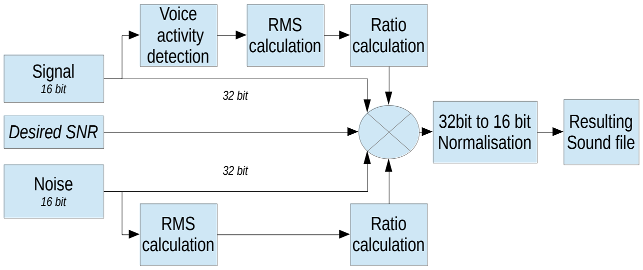


Figure 4. Noise addition scheme.

B. *Results*

We tested the possibility of assessing the effect of noise presence on the quality of Automatic Speech Recognition (ASR) using both synthetic and real speech. The reference sample of real speech consisted of the recording of 100 phonetically rich Slovak sentences collected for another project [14]. The same sentences were also generated using HMM TTS described in Section III above. Noise of four types (white, pink, brown, and traffic noise) and five SNR levels (-30 dB, -25 dB, -20 dB, -15 dB a -10 dB) were mixed with the all the original and synthesized speech signals. The resulting sentences then served as an input into our basic ASR system for Slovak [15]. The quality of recognition was assessed with a standard Word-Error Rate (WER) measure. The results are summarized Table I.

The table shows that brown noise deteriorates the speech signal the least, while the traffic noise distorts speech the most.

The results averaged for the type of noise are shown in Fig. 5. Two observations can be made. First, the degradation of ASR performance is non-linear. While increasing the noise levels between -30 and -20 dB results in rather moderate decrease in ASR performance, the last step produces a sharp decline in ASR performance.

TABLE I. WORD ERROR RATE (WER) RESULTS FOR DIFFERENT TYPES AND LEVELS OF ADDED NOISE

Test signal mixture		**SNR**					
		WER [%]					
Noise	**Signal**	**clean**	**-30 dB**	**-25 dB**	**-20 dB**	**-15 dB**	**-10 dB**
White	Human	11.0%	11.4%	13.9%	18.9%	27.2%	43.7%
	TTS	11.8%	12.1%	12.5%	16.0%	23.7%	48.9%
Pink	Human	11.0%	11.2%	14.1%	20.6%	28.7%	55.5%
	TTS	11.8%	12.0%	14.5%	18.5%	31.2%	65.7%
Brown	Human	11.0%	10.8%	11.0%	11.8%	15.2%	24.3%
	TTS	11.8%	10.6%	11.0%	12.1%	14.3%	21.0%
Traffic	Human	11.0%	14.1%	20.8%	25.6%	36.2%	64.6%
	TTS	11.8%	13.5%	16.6%	24.9%	44.9%	73.2%

Second, the results for human and TTS-produced speech are comparable with high correlation between them.

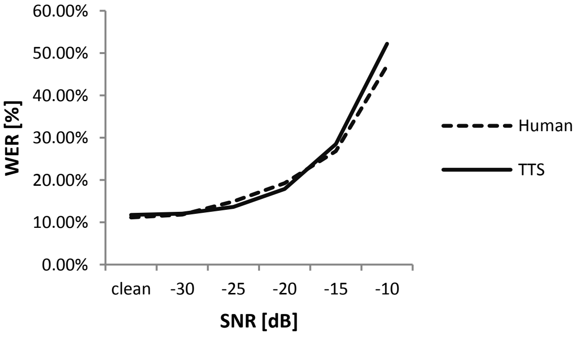


Figure 5. Comparison of Automatic speech recognition (ASR) performance in noise for human and synthesized (TTS) speech based on Word Error Rate (WER) for various levels of Sound to Noise Ratios (SNR).

This is an important observation since it provides a proof of concept that noise effects on these two types of speech result in comparable effects on understandability. This, in turn, will facilitate and accelerate significantly the production of test recordings for the evaluation of ASR systems in adverse acoustic environments.

VI. CONCLUSION

The paper outlined our work on a new system for generating speech samples that are suitable for testing the quality of speech recognizers deployed in adverse acoustic conditions. This system also facilitates parametric studies and experiments with optimizing speech synthesis systems for high intelligibility in noise conditions or with distorting sound effects. The unique functionality of speech effort control allows simulating various vocal modes including shouted speech, Lombard speech, or long-distance speech. The user interface allows for fast online signal generation and the flexibility of the systems allows its implementation in various designing solutions.

ACKNOWLEDGMENT

This publication is the result of the project implementation: Technology research for the management of business processes in heterogeneous distributed systems in real time with the support of multimodal communication, ITMS 26240220064 supported by the Research & Development Operational Program funded by the ERDF.

REFERENCES

[1] A. W. Black and B. Langner, “Improving Speech Synthesis for Noisy Environments,” Speech Synthesis Workshop 7 (SSW7), Kyoto, Japan, 2010, pp. 154-159.

[2] M. Cerňak, ”Unit Selection Speech Synthesis in Noise,” Proc. of ICASSP06, Toulouse, France, May 14-19, 2006, pp. 14-19.

[3] R. Vích, J. Nouza, and M. Vondra, “Automatic Speech Recognition Used for Intelligibility Assessment of Text-to-Speech Systems,” Verbal and Nonverbal Features of Human-Human and Human-Machine Interaction, Springer 2008, pp. 136-148.

[4] R. H. Wilson and W. B. Cates, “A Comparison of Two Wordrecognition Ttasks in Multitalker Babble: Speech Recognition in Noise Test (SPRINT) and Words-in-Noise Test (WIN),” Journal of the American Academy of Audiology, vol. 19, no. 7, 2008, pp. 548-556.

[5] L. Couvreur and C. Couvreur, “Robust Automatic Speech Recognition in Reverberant Environments by Model Selection,” Proc. of the International Workshop on Hands-Free Speech Communication, Kyoto, Japan, 2001, pp. 147-150.

[6] T. Yoshioka et al., “Making Machines Understand us in Reverberant Rooms: Robustness Against Reverberation for Automatic Speech Recognition,” IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 114-126.

[7] J. H. L. Hansen et al., The impact of speech under ’stress’ on Military Speech Technology, NATO RTO-TR-10, AC/323(IST)TP/5 IST/TG-01, 2000.

[8] A. D. Conkie, “Robust Unit Selection System for Speech Synthesis,” Joint Meeting of ASA, EAA, and DAGA, paper 1PSCB_10, Berlin, Germany, 1999.

[9] S. Darjaa et al., “HMM Speech Synthesizer in Slovak,” 7th International Workshop on Grid Computing for Complex Problems (GCCP), Bratislava, Slovakia, Institute of Informatics SAS, 2011, pp. 212-221.

[10] H. Zen et al., ”The HMM-based Speech Synthesis System Version 2.0,” Proc. of ISCA SSW6, Bonn, Germany, 2007, pp. 294-299.

[11] M. Rusko, M .Trnka, and S. Darjaa, “Three Generations of Speech Synthesis Systems in Slovakia,” Proc. of the XI. International Conference SPECOM 2006, St. Petersburg, Russia, 2006, pp. 449-454.

[12] J. C. Junqua, “The Influence of Acoustics on Speech Production: A Noise-induced Stress Phenomenon Known As the Lombard Reflex,” Speech Communication, vol. 20, no 1-2, 1996, pp. 13-22.

[13] M. Rusko, S. Darjaa, M. Trnka, and M. Cerňak, “Expressive Speech Synthesis Database for Emergent Messages and Warnings Generation in Critical Situations,” Language Resources for Public Security Workshop (LRPS 2012) at LREC 2012 Proceedings, Istambul, 2012, pp. 50-53.

[14] O. Jokisch et al., “Multilingual Speech Data Collection for the Assessment of Pronunciation and Prosody in a Language Learning System,” SPECOM´09, 13-th International Conference on Speech and Computer. Editor A. Karpov, Russian Academy of Science, St. Petersburg Institute for Informatics and Automation, State University of Aerospace Instrumentation, 2009, pp. 515-520.

[15] M. Rusko et al., “Slovak automatic transcription and dictation system for the judicial domain,” Human Language Technologies as a Challenge for Computer Science and Linguistics, 5th Language & Technology Conference, Poznań, Fundacja Uniwersytetu Im. A. Miczkiewicza, 2011, pp. 365-369.

Future Illustrative and Participative Urban Planning

Developing Concepts for Co-creation

Virpi Oksman, Antti Väätänen, Mari Ylikauppila
VTT Technology Centre of Finland
Tampere, Finland
e-mails: {virpi.oksman, antti.vaatanen, mari.ylikauppila}@vtt.fi

Abstract – **The starting point of this paper is to develop and experiment with new participatory web-based design services to visualize future urban environments with mixed reality and other content technologies. We have created new visualizations and virtual environments by mixing panoramic imaging and architectural drawings of future urban plans. In order to involve citizens in urban planning projects, we have also implemented user-centred interactions such as questionnaires and commenting tools. In this paper, we discuss how new visual web-based service concepts using mixed reality technologies can be used for future participatory urban planning. To ensure political, economic and social relevance of the developed urban planning service concepts, we have conducted an interview study that clarifies qualitatively, how political decision-makers and other stakeholders perceive the new digital concepts. In addition, we have piloted our participative urban planning demo with users. In the political decision-making processes, the new tools were expected to bring certainty and eliminate uncertainty. New participatory design tools for urban planning should also be efficient at collecting and processing user feedback and other data.**

Keywords-visualization; 3D graphics; urban planning; panoramic imaging; co-creation; participatory design

I. INTRODUCTION

In recent years, many cities and communities have started to pay attention to openness and transparency of decision-making for citizens. For instance, when planning new important urban environments, such as public buildings, energy systems and traffic solutions, different kinds of collaborative workshops are organized for residents to share information on the plans and discuss their impact on the environment. Opening complex urban planning processes and using participatory design or open innovation tools can generate new ideas and change the decision-making to make it more interactive and integrate company representatives and citizens [1][2][3].

In this paper, we discuss how new visual web-based service concepts using mixed reality technologies can be used for participatory urban planning. With mixed reality we refer to the merging of real and virtual to produce new environments and visualizations. Our aim is to develop new mixed reality solutions to visualize future urban environments by, for instance, mixing panoramic imaging and architectural drawings and sketches of future city building projects. With these mixed reality services, we aim to make the plans more visual and understandable to different stakeholders. Our aim is to be able to visualize and discuss the impacts of future building projects and traffic solutions on their environment at the early stages.

Recently several mixed reality technologies including smart phone augmented reality systems have been developed to open up and support stakeholders' participation in urban planning [4][5][6][7][8]. However, mixed reality technologies for urban planning are often developed separately from web-based open innovation and advanced user interaction tools. Our aim is to develop these both under the same service so that up-to-date, visual information should be easy to find and leaving comments would be possible for citizens and other stakeholders. The developed service can be used to promote communication between stakeholders and make decision processes more efficient. By producing easy-to-understand visualizations, it will be possible to view and compare alternative plans and involve citizens and other stakeholders in the planning of the ecology, functionality and quality of their living environments. In addition, we are interested in to find out, what kind of set of participatory concepts support co-creation and stakeholders' participation in urban planning. In what kind of digital environments and public places they should be situated so that users will notice them? What kind of devices and interactive user-interface concepts support participation?

As many new approaches and procedures demand political and social acceptance, in this paper, we will first explore qualitatively how political decision-makers, municipal officials and companies perceive new visual and participatory urban planning service concepts. We interviewed these stakeholders during 2013. The central theme in these interviews was how these new digital tools support decision-making processes and citizen participation in urban planning and how digital urban planning products and services should be developed. In addition, we have conducted preliminary user studies with participation in a small local community in Western Finland, where several environmental urban planning projects are taking place. These projects include, for instance, international airport area development, supplementary construction, green design and planning of noise barriers. Through interviews, queries, demos and case pilots we gained an understanding of how users perceived this kind of participatory mixed reality services in real urban planning projects and how to develop the service further for massive, large-scale participatory projects.

This paper proceeds as follows. In Section II, we discuss the participatory approach to urban planning. In Section III, we describe our research setting. In Section IV, we present the three different co-creation concepts to urban planning which were used in the interviews of political decision makers In Section V, we look over the feedback from political decision makers and other stakeholders. In Section VI, we present our participatory urban planning service demo. Section VII describes citizens' feedback on participatory urban planning demo. Section VIII presents our conclusions regarding the role of participatory services in urban planning and we also discuss our future work.

II. PARTICIPATORY APPROACH TO URBAN PLANNING

In many research fields, such as human-centred design, marketing and service design, the emphasis on user involvement has shifted from treating customers, users and citizens only as passive research objects to taking them into the design process as active co-creators, thinkers and partners. This view has been given a different name and a slightly different emphasis in definitions. Two widely adopted perspectives have been *participatory design* and the *user-centred approach.* Participatory design has often been defined as a shift in attitude from designing for users to one of designing with users. However, it is quite difficult to draw the line between user-centred design processes and participatory experiences. Participatory design is not simply a method or set of methodologies but more of a mind-set and attitude to people. The belief is that all people have something to offer to the design process and that they can be both articulate and creative when given the appropriate tools with which to express themselves. Moreover, participatory design is an approach in which potential end-users have a critical role in the outcome [9][10].

According to service business research, organizations and companies can compete through *co-creation*, innovating value with customers or user communities instead of just doing things for customers [11].

In addition, service business is different from simply providing goods or products. In service business, the value comes, especially, from the ability to act in a manner that is beneficial to the other party. *"Value is subjective and always ultimately determined by the beneficiary, who in turn is always a co-creator of value."* [11]

In urban planning, reaching out and engaging citizens and other stakeholders in making plans is a cornerstone of good practice. Moreover, the collaboration between all the stakeholders in the process – citizens, planners and decision-makers – is the context in which plans are made. The final outcome and plans emerge from the interaction between all the involved stakeholders. The *Open innovation* approach, which comes from the business strategy field, can therefore add valuable insights into service development and enrich a company's or organization's knowledge [12]. In addition, the *crowdsourcing* method can be used with web-based solutions to create the best solution when widespread experimentation and large-scale feedback is needed. Yet, attracting a diverse group of citizen participants can be challenging, since citizen involvement is often a leisure-time activity and competes with other ways of spending time. Developing new visual tools, such as smart-phone augmented reality for public participation in urban planning, can increase users' willingness to participate in urban planning events. At least new AR visualisations can help people to visualise the intention of the design better than with traditional drawn plans [6][4]. Moreover, digitizing services and publishing them on-line makes them more visible to citizens and allows them to participate any time they want. Participatory services should, above all, provide a shared environment for productive, collaborative development [13].

Social media and social applications have been used for open innovation in land use and urban construction projects. A visualized map enables to collect citizens' comments and development proposals. The growing knowledge and power of end users, sustainability requirements and financial restrictions create challenges for traditional urban planning methods. Many industry examples have demonstrated that open innovation and social media extend the traditional data with citizen participation feedback [14]. Web-based public participation and proper technologies can help to involve new groups of citizens in the planning process [15]. Participatory design approach in urban planning can use different set of technologies and methods to create a shared vision of an urban project. For instance a portable lab, called MT –Tent, using Mixed Reality has been used for participatory design on site [3].

All the above-discussed approaches – participatory design, user-centred approach, co-creation, open innovation and crowdsourcing – are relevant and add value to the development of our web-based participatory urban planning design service. However, it is still unclear how critical numbers of representative citizen groups can be encouraged to participate in urban planning and be motivated to make important contributions.

In the next section, we will discuss in more detail our approach to participatory urban planning and how decision-makers and other stakeholders can acquire these new service models and concepts.

III. THE RESEARCH SETTING

The aim of the interviews was to find out how local political decision-makers, municipal officials and companies perceive the need to develop current urban planning methods. The participating political decision-makers were members of environmental or technical boards with a central role in organizing services related to urban planning. The political decision-makers were selected from all the parties presented at these boards. We selected both genders, as well as experienced and new decision-makers who were in their first term on the Board of Governors. We also interviewed five companies, which represented building, architecture and visual Internet-based services.

We used half-structured theme interview method to discuss and collect feedback from new digital visualization and participatory design tool concepts. The interviews took approximately two hours and they were taped and transcript. At the beginning of the interviews we discussed the recent urban planning practices and their challenges.

The central theme in these interviews was how these new digital tools and services should be developed to support decision-making processes and citizen participation in urban planning.

In the interviews, we presented three future urban planning service concepts (described in Section IV). Our aim in the interviews was to place the urban planning concepts in order of importance so that we can choose the kinds of digital urban planning concepts that should be developed in the near future. The services should inspire users and arouse interest in developing better and more versatile living environments. We also aim, through visualizations, to produce better material for decision-making so that it will be possible to view and compare different options.

Moreover, we conducted our first user pilot in a small local community in Western Finland. We wanted to ascertain how to support citizens and other stakeholders in involving them planning of the sustainability and quality of their living environments through digital services. We wanted to find out how our demo service suited this purpose, and how to develop it further, especially trying to understand user values, needs and preferences in participative urban planning. We first conducted a user study in a small village near the highway where a new noise barrier is planned to protect inhabitants from noise pollution. There are only town houses in this area, and residents of the village consisted mainly of families with children and older people. Our aim was to reach residents living near the noise barrier to respond to our inquiry, so we published an online questionnaire link in a municipal community web portal, community Facebook site and in a local newspaper. The query was available over a period of a few weeks in March and in April 2014.

The questionnaire included both multiple choice and qualitative open-ended questions. The survey included basic background information questions, and focused on topics such as clarifying requirements for a future community planning, perceptions on visualisation and participation services, and most preferred places and information channels and devices for utilising a future participatory urban planning service. Users were also asked to try out the web-based pilot service, which mixed panoramic imaging and architectural drawings of the planned noise barrier near their homes. The demo illustrated noise barrier building stages and the area five and twenty years later.

IV. THREE APPROACHES TO PRESENTING FUTURE URBAN PLANS

In the beginning of the interviews, examples of three different ways of demonstrating future urban plans were introduced. The examples helped in figuring out the idea of new visual approaches to community planning and aimed to facilitate feedback and ideas related to the different approaches. Finally, the participants were asked to prioritize the three approaches and state reasons for their preference order.

A. *On-site mixed reality mobile tools*

We aimed to describe possibilities of visualizing urban planning solutions with smartphones and tablet devices. The idea is for users to be able to move around the surroundings under development and see merged virtual 3D objects and a camera view on a handheld device (Figure 1). The virtual building objects will be located in their intended locations. The demonstrated mobile mixed reality tool for architectural sites has been described and evaluated in earlier studies [4][5].

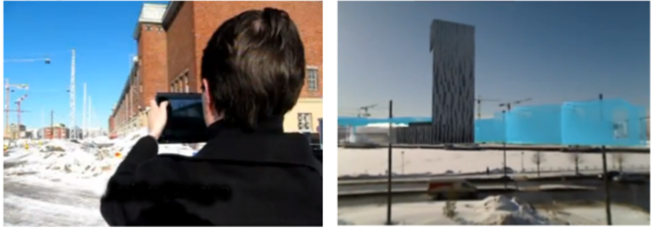


Figure 1. On-site augmented reality solution

B. *Interactive public screens*

The other presented approach was interactive public screens with mixed reality features (Figure 2). The screen shows areas under development, and new digital visualizations are embedded into the views. Users can manipulate the views and community plan options using their gestures or the touch screen input method. Gesture recognition would be implemented with the help of depth camera sensors. This kind of public screens can be located next to the area, in shopping centres or in municipal office buildings.

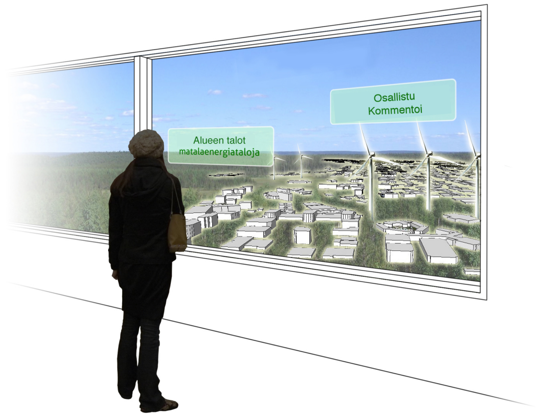


Figure 2. Concept of an interactive public screen with AR features

C. *Off-site interactive design tables*

Thirdly, the users can explore urban planning solutions using interactive and multiuser design tables (Figure 3). The tables can be a combination of tangible objects or 3D printed building models, projected information and camera recognition systems. The users are able to browse different urban planning options or manipulate objects on a table, and they can receive more information using, e.g., pointing, touching or gestures.

Dalsgaard and Halskov have developed and studied a tangible 3D table-top system in which physical objects on a table can be recognized [7]. The same kind of table-top systems in urban planning include the Spatial Design Table and the Bionicle Table [8][16]. They both enable 3D visualizations showing how different buildings look in their environments. The user moves and indicates building options using AR markers on the table.

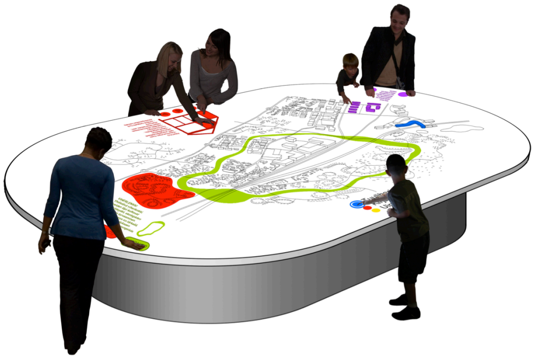


Figure 3. Visualization on an interactive design table

V. FEEDBACK FROM POLITICAL DECISION-MAKERS

Altogether, 13 political decision-makers and municipal officials took part in the interviews. All the interviewed decision-makers attached great importance to developing methods to illustrate urban plans and support citizen participation in future urban planning. According to the interviews, the recent urban planning approaches could be improved by paying more attention to the availability of information and inclusion of citizens at the right time in the urban planning processes.

Recent participatory methods in urban planning projects cannot be applied to all citizens as such. For instance, public workshops are connected to a certain time and place, and busy working families and younger age groups, in particular, are often left out. Young segments have showed more interest in on-line surveys, but reaching younger age groups and getting them to become actively involved and to participate in urban planning presents a clear challenge.

The interviewed politicians perceived the development of information processes and increasing awareness of on-going projects and statements as especially important so that all citizens would have the opportunity to obtain up-to-date information on important projects if they wanted to. The information on on-going urban planning projects is usually available on the city net portal or in paper format at the municipal office. However, not all citizens are capable of acquiring the necessary information. Participation requires personal interest, activity and information seeking to be possible. Versatile information channels can support information seeking, sharpen communication and lower the threshold for participation.

Presenting alternative plans through visualizations would also be important, and plans that are too detailed and complete should be avoided. Overall, the proliferation of new technologies in participatory urban planning is affected by, among other things, the maturity of the technological solutions, implementation expenses, acceptability and ease of use.

When citizens and other stakeholders are asked for feedback and comments on urban plans through, for instance, on-line surveys, the response material needs to be processed, analysed and reported carefully to urban planning officers, planners, decision-makers and citizens. Through on-line surveys, it is easy to access large populations. However, analysing large-scale survey material takes time, work and recourses. Processing large-scale material also demands good and suitable tools.

Of the three presented technology approaches, the decision-makers appeared to prioritize lightweight, web-based mobile solutions, which are suited to illustrating different alternative options in urban planning. Other presented solutions, such as the interactive design table and public screens, were also seen as viable in the long run. They were seen as suitable for large urban planning projects and as tools for both decision-makers and citizens. Public screens were seen as effective attention grabbers and information channels: they were considered a good way of spreading knowledge of urban planning projects. However, screens were seen as less suitable for collecting feedback and ideas from the general public. It was assumed that people would be hesitant to use a technical device that was for public use. The actual participation and feedback would happen via a personal mobile or other personal device, or in a more closed facility organized by the city or community. User interfaces that recognize gestures were seen as better suited to public spaces than touch screens. The interactive design table was thought suitable for concretizing urban plans by decision-makers and active citizens who wanted to participate in urban planning.

VI. PARTICIPATORY URBAN VISUALIZATION SERVICE: ILCO CITIES

Based on the interview results and concept design outcomes, two demonstrators were developed. These ILCO Cities demonstrations illustrated different community plans.

Encouraging residents and other stakeholders to participate in urban planning is a fruitful approach in many ways. When the architectural sketches are presented in illustrative and visual ways, the projects are especially likely to proceed fluently with fewer complaints to slow down the processes. When possible problems can be detected at an early stage and costly changes avoided, the final outcome of the urban planning is usually of better quality.

The stakeholders in urban planning and the users of the new community planning approach can be categorized into three groups: 1) decision-makers, 2) companies, and 3)

ISBN: 978-1-61208-342-1

citizens (Figure 4). Companies are counted as actors in the building industry, e.g., architectural and construction business. Local politicians and city officials are decision-makers who prepare initial plans and processes of community development programmes and activities. Citizens can be called end-users of local community plans. They live and work in the planned environments. All these groups should have transparent, real-time and equal communication of commonly shared living environment design solutions. We have divided the process into three main areas that should be taken into account when creating and implementing the future community plan. The first phase is visualization. The citizens and decision-makers are not usually urban planning business professionals and do not have the capabilities to perceive 2D architectural community or building plans and conceive their effects on the surroundings. On the other hand, building industry actors need to have impressive, cost-effective and easy-to-use tools to represent their plans or ideas. The participation step should offer equal and real-time ways to analyse, prioritize and comment on plans. Citizens, in particular, need to be encouraged and motivated to give their feedback, which requires open information sharing via commonly used media channels and technologies. The final outcome, influence, meaning democratic decisions and diverse possibilities to affect plans, can be achieved. It may enable or demand changes to existing community plan processes. In an ideal case, this kind of advanced operational model can streamline ways to consider, effectively and transparently, the needs and feedback of different stakeholders, and in this case officials proceed faster in municipal decision phases.

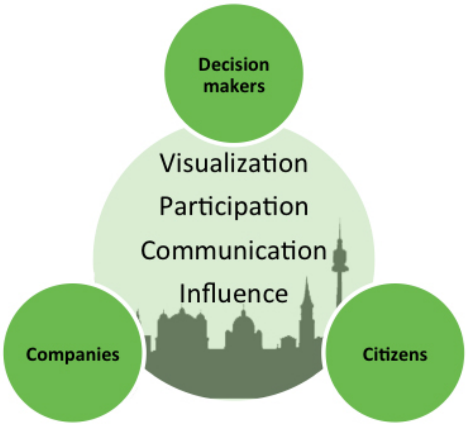


Figure 4. The ILCO Cities service model shows collaboration between different stakeholders in participatory urban planning processes

Two ILCO Cities service demonstrators were developed to concretize the presented service concept. Both demonstrators are related to real on-going community planning cases in two municipalities in Finland. Technically, the demos run on web browsers of different devices such as tablet devices and PCs.

First, the ILCO Cities demo included options for a new congregation building in a small municipality, Lempäälä in Western Finland (Figure 5). The original building was a flat single-layered white building from the 70s, and the future options included two modern, higher and multi-layered architectural plans. The users were able to investigate the following alternatives:

- Current building
- Building model with dark wooden walls and a copper roof
- Building model with white plastered walls and a painted roof

The building models are integrated into the panoramic images of the surroundings close to the congregation building and the user can change view freely in panoramic images. Overall, three panoramic images were taken and used in the demo.

The users change building options by selecting the required model from the menu on the right (see Figure 5). Panoramic image viewpoints can be changed using the eye icons on the screen. Questionnaires are available on the left side, and the user can show or hide the on-line questionnaire. The existing congregation building is architecturally protected, and ILCO Cities aims to generate a discussion on alternatives for the current situation.

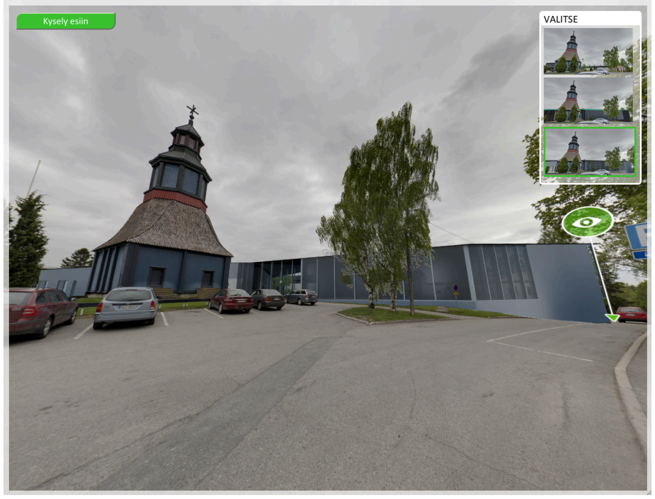


Figure 5. ILCO Cities building options visualization

The second, ILCO Cities demo visualized noise barrier plans between a highway and a field in Pirkkala municipality in Western Finland (Figure 6). The main goal of the demo is to illustrate the following phases of the upcoming work:

- Recent surroundings
- Drawings of noise barriers
- Computer graphics of ready-made noise barriers
- Situation after twenty years when plants such as trees have grown up

The panoramic views were taken from four locations around the planned noise barrier, and the user can change the viewpoints.

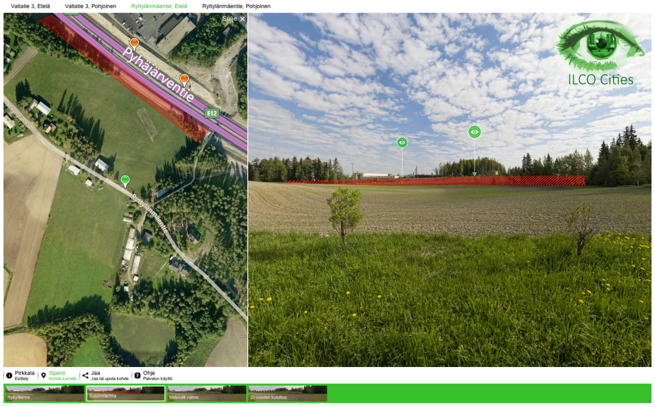


Figure 6. ILCO Cities noise barrier plans illustration

Questionnaires, a map and extra information on the noise barrier can be seen on the left side of the user interface. The user can also see the viewpoints and noise barrier drawings on the map.

VII. CITIZENS' FEEDBACK ON PARTICIPATORY URBAN PLANNING DEMO

In all 25 respondents (12 males, 13 females) completed the questionnaire. Most of them belonged to the age group from 35 to 44 year olds. Their attitudes towards environment-related development activities were mainly very positive.

Respondents were mainly interested in the projects that are linked to their neighbourhood area, somehow reflect their everyday lives or projects that are supposed to have large, revolutionary influences not only geographical but also at the societal level. At present the information of ongoing projects is sought mainly from municipal's websites (76%) and from the local newspaper (80%), which both are listed as municipal official communication channels. Only three out of 25 have visited an official bureau or participated in an event organised by municipal to inform citizens about new projects.

The results of the survey were very much in line with the issues discussed with decision makers. In open-ended questions urban planning information was complained of as being difficult to find, and the participation process is perceived as being too complex. Opportunities to interact and be heard were claimed to be very challenging. It was even said that citizens are given an opportunity to give their feedback, but that feedback is rarely really taken into consideration. Respondents demanded involvement at an earlier phase of the planning process, more alternative solutions to be compared, clear timetables and information on how the process is progressing.

General attitude toward presented demonstration was very positive. The service was found to be interesting, useful, easy to use, and it was thought to bring something new into urban planning and citizen participation. Despite of positive attitude, a common concern related to new participatory methods used in urban planning was, how the results will be used and will there be a real impact.

Participants were asked how the visualisation service succeeded in visualising the example case. Figure 7 indicates respondents' feedback related to how well future visualization and participation tools are applicable for different municipalities' environmental development domains.

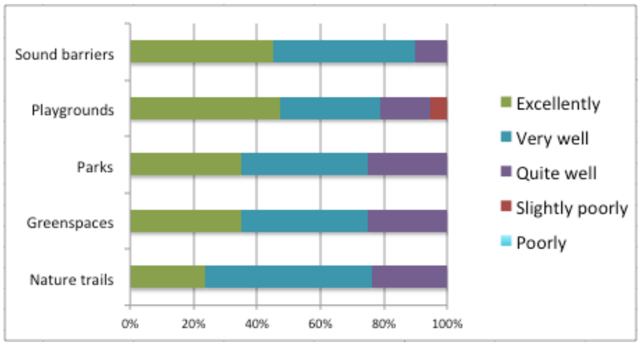


Figure 7. How well does the service suit different environment and sustainability projects?

People are willing to search information and explore the material at home or other private premises. Public spaces such as shopping mall or railway station were seen the most unlikely contexts to take part in urban planning. Lack of privacy and office hours were mentioned as major barriers. Figure 8 shows more detailed, how users would like to have access to participatory urban planning service. They preferred mobile devices as a convenient way of using the services in local environments. However, users reflected that they would be quite unlikely to use it from a municipal service point and also municipal public events and notices in public transport were quite uncertain or unlikely places to access and use these services. In open-ended questions, users reflected that it would be problematic to give their opinions in such public places, if they wanted to maintain their privacy.

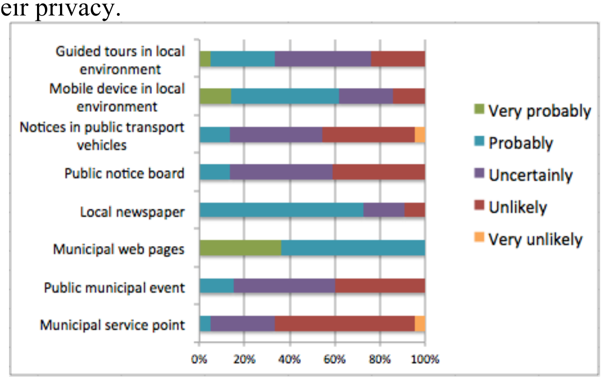


Figure 8. Where would users like to have an access to participatory urban planning service?

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-342-1

VIII. CONCLUSIONS AND FUTURE WORK

In this paper, we have discussed how new visual web-based service concepts using mixed reality technologies can be used for participatory urban planning and co-creation of future living environments with different stakeholders. The new tools were expected to bring certainty and eliminate uncertainty in the decision-making processes. The interviewed decision-makers reflected that they often receive urban plans that are too prepared just to accept or reject. They wanted real team play instead and more open discussions with different stakeholders. Illustrating and visualizing urban plans was thought to enhance the quality of the decision-making materials. The new web-based visualization services were seen as furthering the perception of entireties, complex dimensions, measures and impacts, which were seen as difficult to figure out at present. The new tools were expected to make it possible to illustrate and compare different options and their direct and indirect impacts on the environment. In addition, they would offer users the option to give feedback and share their ideas at any time of the day they wanted. This would be useful, especially in trying to target younger age groups that rarely participate in workshops organized by communities.

A good option to demonstrate future urban plans to different stakeholders is lightweight mobile solutions, which can be taken to different places and situations at any time and used to illustrate alternatives. More demanding approaches, such as interactive design tables could also be useful, especially for large urban planning projects. They were seen as suitable co-creation tools for decision-makers and active citizens. The public display boards were seen as effective marketing methods for new urban plans. However, as long as the system is located in a public place and close to people flows, it will have an effect on participation by shy or privacy-oriented people. The actual participation, e.g., responding to surveys, sharing ideas and feedback, would happen most conveniently with a personal mobile or other personal device. Citizens in general are interested in commenting on and participating in urban planning projects, which are related to their everyday lives and their own neighbourhood. New participatory design tools for urban planning should also be efficient at collecting and processing user feedback and other data. Currently, data processing and analysis of surveys take a large amount of resources. Afterwards, citizens should be informed, for instance, that answering the survey was useful and that their feedback has been taking into consideration in the urban planning. There are currently no proper tools for this.

However, there are many open questions regarding the development of visual, web-based participatory design tools. What kind of things do citizens want to comment on and influence in their living environments? What kind of public projects arouse interest? Are citizens interested in commenting on renewable energy solutions such as the placement of solar panels on public buildings? How can these new tools be used for sustainable urban planning? Should comments and feedback be collected from all possible user segments or mainly or only from the users who are involved in the project in their everyday lives?

In the near future, the ILCO Cities demonstrators will be piloted and used to involve citizens in large-scale urban planning projects. We will proceed to end-user trials in real use situations and compare the way citizens and other stakeholders perceive the demonstration system. The piloting phase will also provide an opportunity to analyse differences between on-site and off-site utilization approaches. Technically, the system could be developed to be more location- and augmented-reality-oriented, which means that the architectural 3D models could be automatically adapted to the real-time camera views. The objects on the screen could also be interactive, and users could add comments or fill out questionnaires by selecting preferred design solutions in the view. Moreover, land use planning projects are often so complex and extensive that using new participatory tools would probably not always impact on their length but on the quality of planning. Digitizing information and increasing on-line channels are the most powerful way to facilitate access to information and make the recent closed urban planning processes more open and participative.

REFERENCES

[1] M. Arnold and V. Barth, “Open Innovation in Urban Energy Systems,” Energy Efficiency, vol 5, no. 3, pp. 351-364, Aug. 2012, doi:10.1007/s12053-011-9142-6.

[2] J. Noujua & L. Soudunsaari, L.,and H.-L. Lentilä, “Boosting Web-based Public Participation in Urban Planning with a Group of Key Stakeholders”, 11th Biennial Participatory Design Conference (PDC '10), IXDA, Dec. 2010, pp. 239-242, ISBN: 978-1-4503-0131-2.

[3] I. Wagner, M. Basile, L. Ehrenstrasser, V. Maquil, J-J. Terrin and M. Wagner, “Supporting Community Engagement in the City: Urban Planning in the MR-Tent”, Fourth International Conference on Communities and Technologies (C&T '09), Penn State University, Jun. 2009, pp. 185-194, ISBN: 978-1-60558-713-4.

[4] T. Olsson. A. Savisalo M. Hakkarainen, and C. Woodward, “User Evaluation of Mobile Augmented Reality in Architecture”, eWork and eBusiness in Architecture, Engineering and Construction, G.Gudnason and R. Scherer, Eds. Taylor & Francis Group, London, pp. 733–740, 2012.

[5] C. Woodward and M. Hakkarainen M, “Mobile Mixed Reality System for Architectural and Construction Site Visualization”, Augmented reality – Some Emerging Application Areas, A. Yeh Ching Nee (Ed.), InTech, Dec. 2011, ISBN 978-953-307-422-1.

[6] M. Allen, H. Regenbrecht, and M. Abbot,”Smart-Phone Augmented Reality for Public Participation in Urban Planning”, The 23rd Australian Computer-Human Interaction Conference (OzCHI ’11), ACM, Dec 2011, pp. 11-20. ISBN: 978-1-4503-1090-1.

[7] P. Dalsgaard and K. Halskov, “Tangible 3D Tabletops: Combining Tabletop Interaction and 3D Projection”, 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design (NordiCHI’12), ACM, Oct. 2012, pp. 109–118, ISBN: 978-1-4503-1482-4.

[8] H. Ishii, E. Ben-Joseph, J. Underkoffler, L. Yeung, D. Chak, Z. Kanji, B. Piper, E., E. Ben-Joseph, and L. Yeung, Z. Kanji, “Augmented Urban Planning Workbench: Overlaying Drawings, Physical Models and Digital Simulation”, International Symposium on Mixed and Augmented Reality (ISMAR’02), IEEE Computer Society, 2002, pp. 203 – 211, doi: 10.1109/ISMAR.2002.1115090.

[9] E. Sanders, “From User-centered to Participatory Design Approaches”, Design and the Social Sciences, Making Connections, J.

Frascara (Ed.), CRC Press, 2002, pp. 1–8, eBook ISBN: 978-0-203-30130-2.

[10] D. Schuler and A. Namioka, “A Participatory Design: Principles and Practices”, Routledge, 1993.

[11] R. Lush, S.Vargo and M. O’Brien, “Competing through Services: Insights from Service-dominant Logic”, Journal of Retailing, Elsevier, vol. 83, nro. 1, 2007, pp. 5–18, doi:10.1016/j.jretai.2006.10.002.

[12] E. Selzer and D. Mahmoudi, “Citizen Participation, Open Innovation and Crowdsourcing: Challenges and Opportunities for Planning”, Journal of Planning Literature, 2012, vol. 28, nro. 1, pp. 3-18, 10.1177/0885412212469112.

[13] C. Skelton, M. Koplin, and V. Cipolla, “Massively Participatory Urban Planning and Design Tools and Process: the Betaville Project”, 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times, ACM, Jun. 2011, pp. 355-358, ISBN: 978-1-4503-0762-8.

[14] J. Porkka, N. Jung, J. Päivänen, P. Jäväjä, and S. Suwal, “Role of Social Media in the Development of Land Use and Building Projects”, EC PPM Conference proceedings, 2012, pp. 847-854.

[15] J. Noujua, “WebMapMedia: A Map-based Web Application for Facilitating Participation in Spatial Planning”, Multimedia Systems, 2010, vol. 16, nro. 1, pp. 3-21, Online ISSN: 1432-1882.

[16] R. Nielsen.J. Fritsch, J., K. Halskov, and M. Brynskov. 2009. “Out of the Box – Exploring the Richness of Children’s Use of an Interactive Table”, 8th International Conference on Interaction Design and Children (IDC), ACM, Jun. 2009, pp. 61-69, ISBN: 978-1-60558-395-2.

Musing: Interactive Didactics for Art Museums and Galleries via Image Processing and Augmented Reality

Providing Contextual Information for Artworks via Consumer-Level Mobile Devices

Gentry Atkinson, Kevin Whiteside, Dan Tamir
Department of Computer Science
Texas State University
San Marcos, TX, USA
{gma23, kjw52, dt19}@txstatet.edu

Grayson Lawrence, Mary Mikel Stump
School of Art and Design
Texas State University
San Marcos, TX, USA
{gl16, mr14}@txstate.edu

Abstract—**Textual didactics, used in museums and galleries, provide access to historical, socio-political, technical, and biographic information about exhibited artworks and artists. These types of didactics are considered to be cost-effective. However, they do not enable the use of audio, video, and Web interface that allow for multiple forms of usage for the museum visitors. We have developed a smartphone application, called *Musing,* for interaction of museum visitors with informational content and enhancement of their museum experience. *Musing* is an augmented reality (AR) application that enables the visitor to capture an artwork with a smartphone camera. Using image processing, the application recognizes the artwork and places graphical user interface objects in the form of Points Of Interest (POIs) onto the image of the artwork displayed on-screen. These POIs provide the visitor with additional didactic information in the form of text overlays, audio, video, and/or Web sites. The *Musing* application, described in this paper, is designed with several performance and efficiency goals, including high reliability and recognition rate, high usability, and significant flexibility. The application is designed to be adaptable to a variety of museums and galleries without requiring special hardware or software. Furthermore, an administrative interface enables museum staff to provide content for the didactic purposes without requiring software development skills.**

Keywords—interactive didactic; museum didactic; virtual meseum; image recognition; augmented reality

I. INTRODUCTION

Museums have historically been tasked with providing access to and educating visitors about artworks. Museum didactics attempt to clarify artworks' meanings by addressing concepts of art, history, geography, politics, and artistic medium/techniques, as well as the lives of artists. For many visitors, however, museum and gallery exhibitions may lack the proper context to allow access points for exhibited works and can leave the "uninitiated viewer" feeling intimidated, "particularly when it comes to interpretation" [1].

In many ways, mobile technologies, such as responsive Web sites and Augmented Reality (AR), present an ideal opportunity to make those personal connections with the visitor, as well as help the visitor make connections to the exhibited objects and/or works of art. As such, the context for the artwork is broadened via interviews, videos, Web sites, source material, art historical influences, and other artworks with shared conceptual frameworks, all of which can be integrated into a mobile application for the museum. Such a personalization of experience through narrative is a highly effective way to expand the context for the work and deepen viewers' connections as they process and integrate the information into their existing world-view [2].

Nevertheless, under the current paradigm, in order to add audio and video to exhibits, museums must rely on proprietary hardware and software. The hardware must be provided by the institution at significant cost both in capital investment and in maintenance. The software used on these devices is often proprietary for the exhibition, reliant on external hardware installed in the gallery, and must be reprogrammed for new exhibitions. While large museums may have the resources to purchase and maintain these systems, smaller community-based museums often do not.

Pedagogical shifts away from passive to active participation are occurring in higher education, as well as in museological practices, and reflect the changing needs of the visitor [3]. An enriched learning environment requires incorporating diverse learning styles including visual/print, visual/picture, auditory, kinesthetic, and verbal/kinesthetic modalities [3].

A. Problem Statement

In order for museums and galleries to fully meet the needs of their visitors, they must incorporate didactic information that embraces diverse learning styles and present multiple types of didactic information.

In order to reach the highest number of museums and their visitors, an interactive didactic system should be designed, which does not rely on proprietary hardware, the installation of external devices in the gallery, or the need to reprogram the system when exhibits are modified or added.

In order to create a system that does not require proprietary hardware, the system should be developed on mobile hardware that many of the museum visitors already possess. This hardware would include classes of smartphones and tablets running on iOS or Android operating systems.

In order to minimize the technical burden on museums, the system should not rely on extra hardware such as Bluetooth or Near Field Communication (NFC) devices.

Finally, image processing and image recognition (IR) algorithms should be used in order to provide the opportunity for the viewers to deepen their connections to artworks and remove the need for external tokens such as Quick Response codes (QR) or number codes to be entered by users.

B. *Hypothesis*

By using a combination of off-the-shelf image recognition algorithms and unmodified consumer-level hardware, the research team will be able to create a system that is fast and accurate enough to be usable in a museum, without the need for proprietary hardware or external tokens. In addition, retrieving exhibition data via a database will allow for a client program that is sufficiently flexible and does not require reprogramming when exhibitions are added or modified.

The proposed interactive didactic system will be designed with a client-server architecture. A database will provide the client application with access to didactic information without the need to permanently store that information on the device. The client application will be programmed for current popular hardware such as a smartphone or a tablet, either owned by the museum visitor or provided in the form of a loaned device.

In order to test the relative success of the application and its acceptance by museum visitors, *Musing* will be deployed in an exhibition at The University Galleries at Texas State University, a three thousand square foot, university-based, contemporary art exhibition venue. Benchmark testing of the application will be conducted in order to determine IR accuracy rate and speed. An exit questionnaire will be given to visitors in order to determine their acceptance of the system and perceptions of system performance and usability.

C. *Proposed Solution*

Musing, a mobile, image recognition and AR application, runs on consumer-based mobile hardware, requires no external tokens or hardware, and does not require reprogramming between exhibits. The application has passed the Apple approval process and is available at [4].

The main contributions of this research is the design, development, and deployment of an end-to-end reliable, usable, and effective AR system that provides a museum visitor with virtual information and provides museum staff with adaptable, cost effective, and easy to maintain virtual museum utility. To date and to the best of our knowledge, this is the only fully functional system that integrates hardware agnostic and software agnostic virtual museum content delivery, and administrative support.

This paper is organized in the following way: Section II provides background in the form of relevant past research performed by this team, with Section III containing a Literature review. The application deployment of *Musing* is outlined in Section IV, followed by deployment results showcased in Section V. Section VI explains the evaluation of results from both benchmark testing and exit questionnaires given to the museum visitors. Lastly, Section VII outlines the conclusions and future research for *Musing*.

II. BACKGROUND

A. *Previous Research*

In 2012, the research team developed a series of responsive Web pages triggered by QR codes used in an exhibition at The University Galleries at the Texas State University [5].

In this pilot program, a QR code was included in the tombstone wall label placed next to artworks in the gallery. These codes, when scanned with reader software on the user's smartphone, presented the visitor with a custom-built Web page for each artwork. These pages provided supplemental didactic information via news articles that pertained to the artwork's subject matter, full artist biographies, video interviews with the artist, photos of the artist's workspace, and links to external Web sites.

During the pilot exhibition, the gallery Web site recorded 23 unique visitors per day with an average time on-page of 3 minutes and 37 seconds. The Web pages that were only accessible by the QR codes were responsible for 16 of the 23 unique daily visitors (69%) and the majority of the time on-page (3 minutes and 33 seconds). For comparison, exhibits installed after the pilot test did not include QR codes. The subsequent exhibit showed a decline in both the number of online visitors (-26%) and the amount of time visitors spent on the gallery Web site (-42.5%). This data indicates that when QR codes are included with the artworks in the gallery, there is an increase in both online traffic and online interaction with the visitor.

The experiment with QR codes in the gallery indicated that visitors would use interactive technologies in the gallery and that they would spend the time necessary to consume the extra content. However, a major drawback of the QR codes was the inability for the museologist to contextually place information within artworks' representation. This ability would allow the administrator to place content exactly where it would be most pertinent to the visitor's view of the artwork. For example, a POI could be visibly placed relative to a specific element of an artwork to provide information about that element's significance. Lastly, QR code reader software is not created specifically for the needs of museums and galleries, as they are designed to work for a wide variety of applications, from advertising to stock keeping.

Following the successful response to the QR code project, it was decided that the next step in the research should be to create an AR system that would allow for information placed within an artwork, designed specifically for the needs of museums and galleries.

III. LITERATURE REVIEW

A literature review showed a number of teams researching the possibility of using AR to augment the information provided by museum didactics. In most of the cases, however, these didactics rely on proprietary hardware, require reprogramming between exhibitions, or installation of external tokens (e.g., Bluetooth, RFID, and

QR) within the museum space. Some work has been done with respect to the challenges of image recognition, but little attention has been paid with regard to integrating hardware/software agnostic image based picture recognition with content delivery.

Bimber et al. have developed a mobile system, named, *PhoneGuide* allowing museum visitors to use mobile phones to detect artworks in a physical museum space [6]. Their method includes image recognition, using the phone's camera, as well as pervasive tracking techniques using a grid of Bluetooth emitters distributed in the space [6]. The reliance on external tokens (e.g., Bluetooth) to assist in the object recognition would require the museum to install new hardware and provide for updates in each gallery space.

Hatala et al. describe a prototype system, called Ec(h)o, developed to provide "spatialized soundscapes" for museum visitors [7]. That is, specialized audio is played for the listeners depending on their position within the museum. The supplied audio is meant to improve the overall experience of the exhibit rather than providing information specific to each artwork.

Jing et al. have developed a mobile augmented reality prototype system which uses image recognition running on specialized hardware to provide additional information on physical images displayed in museums for Personal Museum Tour Guide Applications [8]. The system uses the SIFT recognition algorithm that employs "coarse to fine" recognition to improve the speed of the process [9]. Nevertheless, some users complained of slow processing speed [8].

Blockner et al. developed a prototype system which allows users to create virtual museum tours on a mobile app. The mobile device uses NFC to transmit these tours to projectors positioned within the gallery which display the desired information [10].

Miyashitat et al. have developed an interactive device at the Dai Nippon Printing (DNP) Museum Lab at the Louvre Museum (Paris) for use with an exhibition on Islamic Art. This device used a neural network based system to map content of exhibits and was able to recognize three dimensional objects from a single viewpoint, but also relies on purpose specific hardware which is not available outside the Louvre and requires that Bluetooth enabled hardware be installed in the gallery [11].

Klopfer et al. proposed a "location aware field guide" which operated in a manner similar to *Musing* but it was not adapted to use in a museum [12].

Lee et al. used an ultra-mobile PC, inertia tracker and camera for object recognition [13]. This system did not rely on external devices; instead, it relied on template matching. In this case, a translucent image of the next artwork is placed on the screen, guiding the user to the next artwork to be matched and used to locate the user within the museum space, attempting to estimate the user's location by the last artwork scanned. However, this approach does not provide for an accurate location estimate. Furthermore, this project relied on proprietary hardware supplied by the institution.

Another system that used specialized hardware to provide an augmented reality experience is described in [14]. The system overlays the picture of a physical image displayed on a custom hardware with pertinent information in real-time. The detection of the artwork is accomplished using ultrasound sensors and gyros for pose tracking. The information is then matched to the image using an edge-detection algorithm.

IV. Application Development and Deployment

Musing, developed by an interdisciplinary team that included researchers within Computer Science, Communication Design, and Museology backgrounds, was deployed from October 8th, 2013 through November 14th, 2013, in The University Galleries at Texas State University, for the exhibition, *Eric Zimmerman: West of the Hudson* (example images, scanable by *Musing*, are available in [15]). During the 38-day run of the exhibit, 242 visitors downloaded *Musing*. In addition, 11 visitors borrowed iPod Touch devices provided by the galleries, indicating a high number of visitors used their personal devices. Gallery guest book logs showed that a minimum of 962 visitors attended the exhibit, resulting in 25% of visitors choosing to use *Musing*. This indicates a relatively strong initial acceptance rate of the concept. However, these figures do not account for repeat visitors, visitors who did not sign-in at the front desk, or visitors who shared devices.

A. Pedagogical Design

At the heart of the ideal 21st century museum/gallery experience is what educator and innovator John Dewey referred to over a century ago when he spoke of the importance of interactivity to provide for an enriched learning environment [3]. Such interactivity, and the resulting enrichment, requires providing for diverse learning styles by including visual/print, visual/picture, auditory, and verbal/kinesthetic modalities. These enriched learning environments are comprised of seeing, hearing, and interaction by moving beyond the traditional linear model of communication that provides didactic information via textual labels and gallery talks, to a non-linear model of communication through the provision of individual POI associated with each scanned artwork. Through the visitor's ability to access the POIs contained within *Musing*, the application allows for the creation of an enriched environment in which the visitors can participate in creating context for the works exhibited. The provision of additional information about each work via POIs, positions the visitor as a collaborator in the process of making meaning and serves to engage the visitor with the provided information which solidifies the content knowledge [3]. Meaning is made in a variety of ways and looking at art can begin by seeing the work through several different filters. The individual POI provides an opportunity to show the viewer the works within an art historical, biographical, conceptual, or technical framework. As museums and galleries continue to seek ways in which the visitor's experience can be augmented and expanded, these POIs are an easy way to provide access for visitors to more contextual information for the exhibited works, broadening the exhibitions' theses for the novice viewer, as well as augmenting the meaning for the more initiated viewer. This extends the application's ability to meet the needs of a variety of visitors who access works on a multitude of levels. As such, the broadening of the exhibited works' context via interviews, videos, Web sites, source material, art historical influences, and other art with shared conceptual frameworks allows for a personalization for the visitor through the implied narratives [2]. This is thought to be the most effective way to expand the context for the work and deepen viewers' connections through the exercise and action of gathering the information, resulting in the visitors' "[integration of] the information into their existing world view" [2].

For the novice viewer, whose frame of reference may be lacking in depth to fully make these associations, the POI format is ideal to expand reference points. As these associations and connections deepen, the experience begins to look more familiar, something that can also make looking at art more comfortable. As museologist Marjorie Schwarzer writes, "Today, when the meaning of art is more contested than ever, [technologies] offer visitors the possibility of diverse interpretations" [16]. Schwarzer adds, "The branches of information available on these devices are close in spirit to the multiple ways in which we engage art" [16]. The ability to allow for different levels and a wide range of information, as well as a seemingly endless number of interpretive applications, reflects the diversity of the museum audience, itself [16]. Ultimately, the knowledge and deepened understanding that the POIs facilitate are filtered through the learning and innovation skills of the 21st Century—that of creativity and innovation, communication and collaboration, and cross-disciplinary thinking [2][3][16]. The resulting associations within the gallery setting, moving into the viewers' world, are essential to deepening the understanding of subject matter—a result of the user transferring what he or she already knows and reflecting upon it [3].

Musing's effectiveness comes from the immediacy with which the user can access the POIs content and making information available on demand allows for visitors to move freely within the space, not having to rely upon the preconceived schedule of their guide or any predetermined path.

B. User Interface Design

Musing was designed to employ a client-server architecture that allows museum administrators to upload, remove, and alter content, post-deployment. This is accomplished through an administrative Web interface which feeds the shared database. The application retrieves this content as requested by the user. This approach allows the material provided to the user to be as current as possible. Hence, the application is flexible and not limited to "on board" data, allowing any museum to more closely serve the needs of its visitors. The application relies on an open source library called *OpenCV* for the processing and recognition of images which have been captured by the user.

The User Interface was designed in such a way as to adhere to the Apple Human Interface Guidelines for a tab-bar navigation style application. The application consists of the Exhibitions Screen, Scan Artwork Screen, Artwork View Screen, and Favorites Screen.

C. The Exhibitions Screen and the Artwork View Screen

The Exhibitions Screen, depicted in Figure 1a, consists of a list-view of exhibits that a visitor may visit. The list is organized by "Permanent Exhibits" and "Augmented Reality Exhibits". The Permanent Exhibits are previews of the experience that visitors can expect when using the application in-gallery. They contain artworks that can be viewed outside of the gallery setting (e.g., residence, dorm, etc.). This type of exhibit is included to advertise the application's features, to familiarize the user with how the application works, and encourage users to attend a live exhibition. The AR exhibition section includes exhibits that must be attended in person to gain access to the didactic information for the artworks. This view provides information such as the name of the exhibit, the museum in which the exhibit is located (provided more than one organization uses *Musing*), and a representative image to advertise the exhibition. Figure 1b shows a portion of the "Art View" screen: a captured and identified image along with the overlaid POIs.

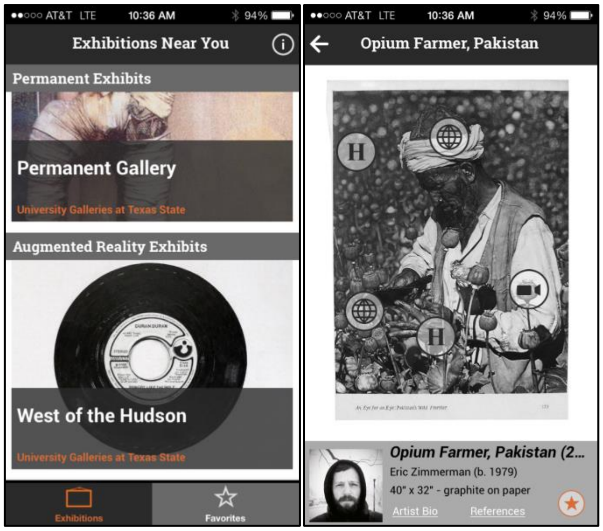

●●○○○ AT&T LTE 10:36 AM 94%
Exhibitions Near You
Permanent Exhibits
Permanent Gallery
University Galleries at Texas State
Augmented Reality Exhibits
West of the Hudson
University Galleries at Texas State
Exhibitions
Favorites
●●○○○ AT&T LTE 10:36 AM 94%
Opium Farmer, Pakistan
H
H
Opium Farmer, Pakistan (2...
Eric Zimmerman (b. 1979)
40" x 32" - graphite on paper
Artist Bio
References

Figure 1: (a-left) Exhibitions Screen, including exhibition selection, and primary navigation; (b-right) A captured and identified image along with the overlaid POIs.

D. Hardware/Software Architecture

Currently, *Musing* runs on iOS-based hardware, such as iPhone, iPod Touch, and iPad. An Android version is under design.

1) Back-end Processing

The back-end (server) application provides two main functionalities. First, it supplies information in the form of reference images and relevant didactic information to the user, enabling its operation inside the gallery or remotely for a permanent exhibition. Second, the back-end is designed to provide an administrator (e.g., a museum staff member) with the capability to edit the contents of an exhibition's didactics within the system. The server, which is shared by the application and the administrative support back-end utility, is used by the gallery administrators to load content into *Musing*.

The back-end was written in PHP and uses standard web-technologies (including HTML, CSS, JavaScript, AJAX, jQuery, and several Open-Source JavaScript libraries) to deliver a user-centric experience. It is designed to allow users unfamiliar with database systems to create, read, update, and delete entries for exhibits from a database stored within the web application's framework. The entries include artworks contained within a chosen exhibit, the associated artists, and curated POIs.

Musing was developed with the intention of packaging within the application as little data as possible. When the user activates *Musing*, it requests an XML document containing a list of available exhibits from the back-end data server. The application parses the XML document and extracts the information into an Exhibit object within the application. Along with the XML document, which contains the names of the exhibits, locations, and id values which the application can use to retrieve data about specific exhibits, the application retrieves a "banner image" for each exhibit, which is displayed in a list for the user to browse.

When the user selects an exhibit from the list, the application passes its id value to a PHP script hosted on the data server. This process is referred to as 'synching'. During synching, the server compiles the pertinent information and returns information in the form of XML file and a set of JPEG images of the gallery artworks to the app. The XML document contains information about each artwork, along with the set of POIs related to the information. The user can tap on POIs to display additional information about the artwork or artist. The images retrieved along with this document are used both for displaying POIs on the Artwork View screen and as references by the image recognition.

As in the case of the exhibit list, the XML document provided by the data server when the application is synched to a particular exhibit is parsed. The extracted information is used to populate painting and POIs within the application for each painting and POIs listed in the database. The images are also incorporated into these objects. Testing has shown that this process of synchronization typically takes approximately 20 seconds, during which time the user is shown a modal progress graphic.

The second functionality of the back-end is to support museum staff in modifying existing exhibition didactics within *Musing* and generating didactics for new exhibitions. This module is still under development. Nevertheless, the following is a description of current and planned functionality.

Artworks and information are added to the *Musing* database using a Web application that can only be accessed by specific museum/gallery staff members and by *Musing* developers.

After selecting an exhibit, the authenticated user is presented with a thumbnail for all of the artworks currently associated with that exhibit. This user is also given the option of adding a new artwork image to the exhibit within the systems. When a new work is added, the user selects an image of the art from local storage on their machine. The image is expected to be cropped such that only the artwork itself and its frame are shown. This greatly improves the recognition performance of *Musing* and creates a more fluid experience for users of the application.

When an image has been selected for a new artwork, the user is directed to a page where information regarding the particular artwork can be entered or edited. This same screen is reached when an existing work of art is selected from the exhibit listing. The user can enter the artwork's title, size, year of creation, medium, and the artist's name. Artists' information is stored and catalogued by the site and details such as year of birth, year of death if applicable, and a link to a biography, can be entered and saved and the user does not need to reenter this information.

Next, the administrative support utility enables the administrator to define and edit POIs for an artwork. This is done using a graphical interface designed with JQuery. The user selects a position on a displayed image of the artwork, chooses what media type that the POI references—along with its associated icon—and the text or URL as appropriate. Users can also alter the position of existing POIs by dragging and dropping them. The user can add and modify exhibits, as well as artists in a manner similar to that described for artworks.

2) Front-end Processing

As noted, *Musing* supports two types of exhibits—permanent and AR. The synching process is the same for both. If the database indicates that an exhibit is permanent, the user is shown a list of artworks available in an exhibit and each may be selected by tapping. This displays the artwork's image with the proper set of overlaid POIs. The second type of exhibit is the AR variety. In this case, the user is given an image detection view rather than a list, which displays a real-time feed from the devices camera over which is laid a graphic of an empty painting frame, along with a button which the user can use to capture a photograph.

During image detection, the users are instructed to position themselves so that a *Musing* enabled artwork fully fills the frame displayed (this is not mandatory, yet it can improve the recognition rate) on the device's screen and to take a picture of the artwork. When this is done and an image is captured, the application compares the captured image to each reference image currently synchronized for the exhibit. If a match can be made, the application proceeds to the Artwork View screen, exactly as it does when the user selects an image in a permanent exhibit. Otherwise, an error message is displayed in a modal dialog. To save in space, the captured image is discarded after being matched or rejected.

From the Artwork View screen, the user has the option of capturing the artwork and its information by making the artwork one of their "Favorites." This is the only condition under which *Musing* locally stores the artwork and its information. This is done by passing the image, POIs data, and artist information to a Favorites Database object that incorporates those values into an array of artwork objects. The data is then written into *Musing's* internal database. The information stored in the favorites array is accessible by the user regardless of whether or not the device is connected to the internet.

Image Processing and Recognition

Musing relies on the Oriented FAST and Rotated BRIEF (ORB) image detection algorithm [9]. The ORB procedure combines the "FAST" key-point detection and "BRIEF" determination of descriptors. Key-points are clusters of pixels within an image which are unusual enough to stand out and to help distinguish a particular image from other images. After identifying a set of key-points within an image, a set of descriptors is calculated for each key-point using BRIEF [17]. This functionality is provided by the *OpenCV* open source computer vision library which is available for use in iOS and Android devices.

Key-point detectors frequently rely on finding "corners" and "edges" within images since image boundaries often create distinguishable pairings of shade and color [17]. By definition, ORB is translation invariant. Additional operations are performed to compensate for rotation and scaling [9].

In the training stage, BRIEF employs binary comparisons between pixels in a smoothed image [17]. This algorithm takes a relatively large set of key-points—often as many as 500—and builds a classification tree for the set. The tree serves as an image "signature" used to measure similarities between images. Alternatively, under the approach used in this research, one can employ the results of the BRIEF stage using the *k* nearest neighbors (kNN) and one-to-one and onto mapping (bijection) test approach.

Following the synching process, users can point their device at an artwork in the gallery and capture its image. This image is processed using ORB and then compared to each of the reference images which were downloaded at sync time. Each reference image is processed to determine its key-points / descriptors at the time of comparison and this information is recalculated for each comparison. *Musing* employs the kNN and bijection approach to the key-points. Each key-point in a captured image is compared to each other in the reference image. A small set of matching key-points in the reference image is found for each key-point in the captured image. The goal is to find a maximal, high reliability, bijection between a subset of the key-points in a reference image and a subset of the key-points in the captured image. Hence, if any key-point in the reference image matches more than one key-point in the captured image with equal reliability, then *Musing* dismisses that match. The literature has suggested 0.65 as a reliability threshold and as the best threshold ratio for selecting one match as superior to the other [18]. The kNN is done twice, creating a set of directional matches that compares the reference image to the photograph taken and vice-versa. Then both sets are compared, dismissing any match that is not bidirectional. If a significant number of bidirectional matches is identified, the images are considered a match. *Musing* currently uses a threshold of 4 bidirectional matches as the minimum subset size.

When *Musing* has determined that a captured image matches a reference image, the reference image is displayed on screen along with an overlay of POIs.

The following is a description of the applied image recognition algorithm, starting with the captured image and the first reference image.

Step One: Captured Image Key-point Calculation - Find the key-points for the captured image using the FAST method [9]. This method checks a ring around each pixel and compares their intensities. It returns the point as a key-point if the gray level of a number of pixels within the ring is sufficiently higher or lower than the nucleus pixel itself.

Step Two: Captured Image Descriptor Calculation - BRIEF is used to take a patch of pixels surrounding a key-point and uses binary intensity thresholds to create a 256-bit binary vector describing the area around the key-point [9].

Steps Three & Four: Reference Key-points and Descriptors - Steps one and two are repeated for the reference image.

Step Five-A: Descriptor Matching (Captured to Reference) - A kNN matching of the Hamming Distances of each descriptor in the captured image to its K nearest neighbors in the reference image is performed. The two best matches for each key-point are retained.

Step Five-B: Descriptor Matching (Reference to Captured) - Step Five-A is applied with the roles of the captured and reference image reversed.

Step Six-A: Ratio-Test (Captured to Reference) - This step discards every match identified for the captured image where the best match and second-best match have similar Hamming distances. This produces a one-to-one match.

Step Six-B: Ratio-Test (Reference to Captured) - Weeding, using the same criteria as in step Six-A is performed on any match from the set of matches identified for the reference image.

Step Seven: Symmetry Cross-Check Test - The Symmetry cross-check test returns only the pairs of matches that are found from the captured image to the reference image and from the reference image to the captured image. This process enables keeping only the strongest symmetric correspondences and maintaining a bijection.

Step Eight: Output if Found - If four or more matches remain after the weeding performed by the ratio tests and symmetry test, the procedure retains the identity of the reference image and returns to step three for the next reference image (if such an image is available). The procedure keeps track of the identity of the image that produced the largest number of matches and outputs its id. If all reference images have been tested and no match has been found, then a message “Image Not Found” along with instructions to the user on how to improve the possibility of match are displayed.

Figure 2 illustrates the process performed in steps 5 to 7.

E. *Design of Testing Instruments*

Testing instruments consisted of quantitative benchmark testing and a qualitative user perception exit questionnaire.

As a part of the quantitative testing, each reference and captured image has been processed to generate 500 identifying key-points in each of 60 total images. The 60 images consist of: ten reference images $(R_1 - R_{10})$ and ten images that served as captured images$(P_1 - P_{10})$. Each of the captured images was captured four additional times for a total of five capturing per image. The first time was with maximum alignment to the reference images the rest of the four where taken with increasing rotation translation and scaling (due to different distance). The maximal rotation was 40 degrees.

The procedure described above was applied to the ten reference images and fifty captured images. A threshold of 0.3% over the percent of matching key-points, which was empirically identified as the most suitable threshold was used by the program and applied to the matching results.

For the qualitative testing, we have used a 23-question exit questionnaire designed to capture feedback from in-gallery users. The questions were written to determine the user’s acceptance of the application, their perceptions of application performance, enjoyment of the application, as well as pedagogical concerns.

V. Deployment Results

A. *Technical Results (Internal Testing)*

Figure 3 shows a heat-map of the results of this experiment. The figure shows a recognition rate of 96.4% with 0% error of type-1 (false positive) and 3.3% error of type-2 (false negative) obtained with $P_{(1,4)}$ and $P_{(1,5)}$. We have found however, that with rotation of more than 45 degrees there were numerous false negatives; but, still 0% of false positive error.

The testing has shown that *Musing* recognizes images with near perfect reliability under ideal conditions, that is, when a user is directly in front of the artwork, has positioned the artwork correctly within the image capture frame, and is not holding the device at an angle. Nevertheless, excessive rotation of the camera while capturing an image diminishes reliability. Our testing indicates that *Musing* recognizes images at a 45 degree rotation with 90% reliability and a 90 degree rotation with 84% reliability. The application performance degrades when the user stands off of the center line when photographing a piece of art, producing a skewed image. A slight deviation from the center (approximately 15 degrees) produced no noticeable change in testing but at greater values (approximately 45 degrees) the system produces 40% true positives and 60% false negatives. As far as can be determined, in the field-deployment testing, the system did not generate false positive results. Furthermore, the user surveys have indicated that the application did not produce a false positive error in use. Additionally, if the user stands too far from the artwork to properly fill the capture frame the reliability has suffered as well, with the reliability rate dropping to 48% at approximately twice the recommended distance. User surveys indicate that the application’s reliability was sufficient to produce a positive experience for most users.

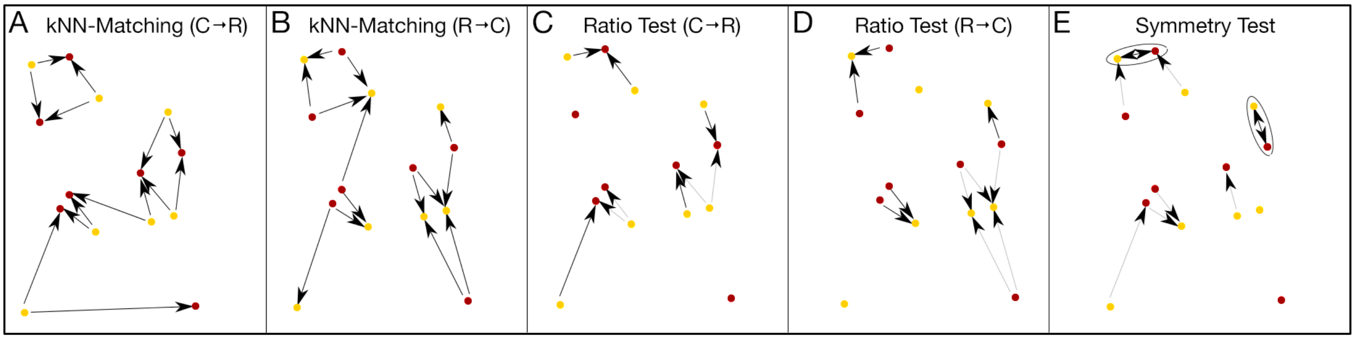


Figure 2: (A) and (B) kNN matching ($k = 2$); (C) and (D) Descriptor matching - the process discards matches with similar quality (Hamming distance) and retains the best match for distinctive matches; (E) Symmetry cross checking – only bidirectional matches are retained.

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-342-1

Testing performed to evaluate the processing time revealed that with 10 reference images, the application was able to compare and either display or reject an image in approximately 3.3 seconds on a stock iPod Touch-5. Again, user surveys indicate that this was sufficient to produce a positive experience for most users.

B. Exit Questionnaire Results with Live Users

Of the pertinent questions, 83.6% responded that *Musing* was able to recognize the artwork "every time" or "most of the time." 77.5% considered *Musing* to be quick and responsive. 87.7% considered *Musing* enjoyable to use and 93.8% wishing to see *Musing* in a future exhibit.

VI. Results Evaluation

The deployment results show high recognition accuracy and relatively short synching/recognition delay time, therefore the functionality of the entire system has been verified. The application has passed the Apple approval process and is available for download [4].

Formal user feedback obtained via questionnaire was consistent with our evaluation of the system and with informal feedback. Visitor responses to *Musing* characterized the application as informative and usable. Their perception of precision and timing was favorable and overall they have commended the system and expressed interest in its further use. Informal feedback from users, including staff members associated with other museums and galleries, was overwhelmingly positive.

VII. Conclusions and Future Research

We have designed, implemented, and deployed a usable mobile application that facilitates an enriched museum visitor experience via AR using interactive didactics. Per our assessment, the application has achieved its stated goals and has shown that the research hypothesis is valid.

The field testing via the exhibition shows that *Musing* can be used on non-proprietary smartphone hardware and provide visitors with didactic information, without the need for external tokens and reprogramming for information changes. This enables reduced reliance on loaner hardware. The implication of such is that the ease of in-gallery application of the technology may allow for higher levels of adoption by individual institutions.

A. Future Research

The University Galleries will be hosting another exhibition deploying *Musing* in the first quarter of 2014. This will provide an opportunity to further asses the capabilities of *Musing*—in specific, several capabilities that have been designed after the first deployment, including the administrative support part of the back-end of the system. This administrative site will allow the application to be deployed in independent galleries and museums by middle to late 2014.

Future enhancements to the *Musing* smartphone application (client) will include abilities for users to share images and didactics via social media such as *FaceBook* and Twitter, as well as the ability to comment on artworks within the application. Additionally, there are plans to complete a port of the current iOS-based implementation to the Android environment.

Other plans for future activities include expanding the image processing capabilities by further improving recognition accuracy, resilience, and time performance. Lastly, we plan to investigate the integration of algorithms for recognition of 3-D objects using the smartphone/tablet camera.

ACKNOWLEDGEMENT

The research team would like to thank Texas State University's Research Enhancement Grant program for providing the initial funding for this research. In addition, continued project funding was provided by the office of the Vice President of Research, the Dean of the College of Fine Arts and Communication, the Dean of the College of Science and Engineering, the Director of the School of Art and Design, and the Chair of the Department of Computer Science of Texas State University.

References

[1] S. Sayre, "Assuring the Successful Integration of Multimedia Technology in an Art Museum Environment," in S. Thomas and A. Mintz (Eds.), The Virtual And The Real: Media In The Museum, Washington, D.C., 1998, pp. 1-10.

[2] K. Morrissey and D. Worts, "A Place For The Muses? Negotiating The Role Of Technology In Museums," in S. Thomas and A. Mintz (Eds.), The Virtual And The Real: Media In The Museum, Washington, D.C., 1998, pp. 147-171.

[3] T. C. Clapper, "The Enriched Environment: Making Multiple Connections" in The Academic Leadership Journal, 8(4), 2010, pp. 1-2.

[4] *Musing*, a photo recognition application that allows users to scan artwork at participating museums and art galleries to learn more about the work, Apple Store, https://itunes.apple.com/us/app/musing/id694382407?ls=1&mt=8, [retrieved March 2014.]

[5] G. Lawrence and M. Stump, "Connecting Physical and Digital Worlds. A Case Study of Quick Response Codes and Social Media in a Gallery Setting," The International Journal of Design in Society, 6(3), 2013, pp. 79-95.

[6] O. Bimber and E. Bruns, "PhoneGuide: Adaptive Image Classification for Mobile Museum Guidance," IEEE International Symposium on Ubiquitous Virtual Reality, Jeju, South Korea, 2011, pp.1-4.

[7] M. Hatala, L. Kalantari, R. Wakkary, and K. Newby, "Ontology And Rule Based Retrieval Of Sound Objects In Augmented Audio Reality System For Museum Visitors." *ACM symposium on Applied computing*, New York, NY, 2004, pp. 1045-1050.

[8] C. Jing, G. Junwei, and W. Yongtian, "Mobile Augmented REality System For Personal Museum Tour Guide Applications", IET Wireless and Mobile Computing, Shanghai, China, 2011, pp. 262 – 265.

[9] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: an Efficient Alternative to SIFT or SURF" IEEE

International Conference on Computer Vision, Barcelona, Spain, 2011, pp. 2564-2571.

[10] M. Blöckner, S. Danti, J. Forrai, G. Broll, and A. De Luca, “Please Touch the Exhibits!: Using NFC-based Interaction for Exploring a Museum,” *International Conference on Human-Computer Interaction with Mobile Devices and Services*, New York, 2011, Article 71, pp. 1-2.

[11] T. Miyashitat, et al., “An Augmented REality Museum Guide, in IEEE International Symposium on Mixed and Augmented Reality, Cambridge, 2008, pp. 103 – 106.

[12] E. Klopfer and K. Squire, “Environmental Detectives—The Development of an Augmented Reality Platform for Environmental Simulations,” in Educational Technology Research and Development, 56(2), 2008, pp.203-228.

[13] D. Lee, and J. Park, “Augmented Reality based Museum Guidance System for Selective Viewings,” IEEE Workshop on Digital Media and its Application in Museum & Heritage, 2007, Chongign, China, pp. 379-382.

[14] J. Oh et al., “Efficient Mobile Museum Guidance System Using Augmented Reality,” IEEE International Symposium on Consumer Electronics, Vilamoura, Portugal, 2008, pp.1,4, 14-16.

[15] *Eric Zimmerman: West of the Hudson*, example images, scanable by *Musing*, http://www.musingapp.com/test_images/, [retrieved March 2014].

[16] M. Schwarzer, “Art & Gadgetry: The Future of the Museum Visit”, Museum News. http://www.aam-us.org/pubs/mn/MN_JA01_ArtGadgetry.cfm, [retrieved, March, 2014].

[17] M. Calonder et al., “BRIEF: Computing a Local Binary Descriptor Very Fast,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(7), 2012, pp. 1281-1298.

[18] E. Rosten, R. Porter, and T. Drummond, “Faster and Better: A Machine Learning Approach to Corner Detection,” Ieee Transactions On Pattern Analysis And Machine Intelligence, 32(1), 2010, pp. 105-119.

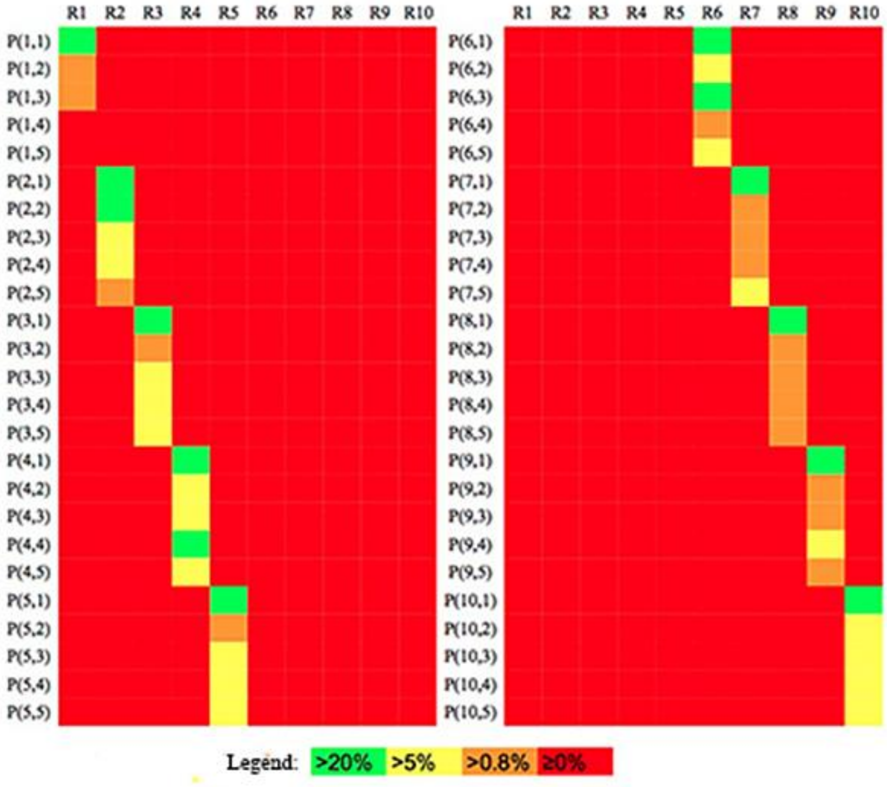


Figure 3: A heat-map of the results of the image matching experiment.

Grade Conversion Model for Joint E-Learning Courses

Jurgita Lieponiene
Dept. of Technological Sciences, Panevezys College,
Panevezys, Lithuania
e-mail: jurgita.lieponiene@gmail.com

Regina Kulvietiene
Dept. of Information Technologies, Vilnius Gediminas Technical University,
Vilnius, Lithuania
e-mail: regina.kulvietiene@gama.vtu.lt

Danguole Rutkauskiene
Dept. of Multimedia Engineering, Kaunas University of Technology,
Kaunas, Lithuania
e-mail: danguole.rutkauskiene@ktu.lt

***Abstract*—Whereas higher education flows into the international study environment, the development of e-Learning takes on higher significance. E-Learning is becoming a means to implement virtual student mobility. Virtual student mobility can be realized by giving join e-Learning courses. Join e-Learning courses are e-Learning courses that are delivered for students from different countries. New tendencies of e-Learning development put forward new demands to e-Learning systems. The application of e-Learning systems for giving joint e-Learning courses includes the support of multilingual user's interfaces, the possibilities of the giving a course content in several languages, and the functions of the grades conversion into the grading scale of student's institution. The article evaluates the application of e-Learning systems for giving joint e-Learning courses and presents a newly created model of the grades conversion from one grading scale to another. The article also describes the integration of a new created model into *Moodle* system.**

Keywords-e-Learning; e-Learning system; virtual student mobility; joint e-Learning courses; grades conversion; plugin of e-Learning system.

I. INTRODUCTION

The priorities of the improvement of the common European higher education environment within the upcoming decade are defined in the communiqué 'The Bologna Process 2020 - The European Higher Education Area in the new decade' [19]. Among them, we enumerate the growth of study accessibility for all social groups, orientation towards student's needs and expectations, the encouragement of students and teachers' mobility, the development of international honesty, and learning from the cradle to the grave.

E-Learning is a means to implement the regulation of Bologna process. The researches have proved that e-Learning makes studying more attractive, allows improving study quality not only in technological, but also in pedagogical terms. E-Learning is a means to implement the mobility of virtual students, to expand united study programs. According to Euler et al. [10], e-Learning is a tool that facilitates the servicing of a new market; it offers the potential to enhance the programme profile of a given university to include other services, as well as being used for the enhancement of university teaching and the implementation of internationalization in education. According to Banks [3], e-Learning is seen as part of globalization to build capacity in 'borderless' education and distance learning, thus improving the competitiveness and marketization of higher education and impact on international cooperation and student mobility.

The changes in higher education environment put forward new demands to e-Learning systems. Whereas higher education is internationalized, e-Learning systems should correspond to new tendencies. They should be adapted to give joint e-Learning courses and should correspond to the needs of students from different countries. E-Learning systems should include the support of multilingual user's interface, the possibilities of the giving a course content in several languages, and the functions for grade conversion into the grading scale of student's institution.

The purpose of the article is to review the application of e-Learning systems for the giving joint e-Learning courses, to present a new created model of the converting of grades from one grading scale to another, and to describe the integration of a newly created model into *Moodle* system [16].

This paper is organized as follows. Section 2 discusses related works. Section 3 analyses the application of e-Learning systems for the giving of joint e-Learning courses. Section 4 presents a new created model of the converting of grades from one grading scale to another. Section 5 describes the integration of a new created model into *Moodle* system. Finally, Section 6 presents our conclusions.

II. RELATED WORKS

The research interest in the internationalization of e-Learning systems constantly increasing. Englisch et al. [9] emphasizes that due to internalization in university study programs more and more multilingual study courses are released. Multilingual support is very important because of globalization [7]. To extend e-Learning systems on a huge number of suppliers and users, multilingual content is necessary [9]. Most of the current available e-Learning systems have an individual implementation of multilingualism [9]. Some of them offer multilingualism only for system text, but not for the user content [9].

Englisch et al. [9] presents a general approach for handling of multilingual content in e-Learning systems.

Denev et al. [7] analyses multilingual support for e-Learning systems and presents multilingual e-Learning solution. Hillier [13] analyses the problems of translation, presents the model for multilingual website.

According to Chen [6], learners' cultural perceptions and experiences influence their online collaboration and communication behavior. When learning communities transcend nations and cultures, this potential influence must be taken into considerations in the design of online courses for cross-cultural collaborative online learning [6]. Mirza et al. [15] emphasizes that the barriers associated with the cultural differences in learning environments become more and more important with the increasing globalization of education. According to Blanchard et al. [5], if the content in a global e-Learning activity is not adapted in function of the culture, there are risks that learners of different culture background consider the same concept in different manners.

Chen [6] presents the design of a cross-cultural e-learning 2.0 environment, which fosters a learning community and facilitates collaborative learning. Blanchard et al. [5] creates a new kind of system called Culturally AWAre System that is centred on Culturally Intelligent Agents, i.e. agents that are able to understand and adapt to cultural specificities of learners. Edmundson [8] presents the cultural adaptation process model as a preliminary guideline for adapting e-learning courses for other cultures.

The performed analysis of literature has shown that scientists analyse the issues, relating to the application of e-learning systems to different cultural and language environments however the aspects of internalization of e-learning systems, evaluating the differences between studies results assessment systems are not studied.

III. THE APPLICATION OF E-LEARNING SYSTEMS FOR THE GIVING OF JOINT E-LEARNING COURSES

Analyzing the application of e-Learning systems for the giving of joint e-Learning courses three open-code e-Learning systems used in Lithuanian higher schools were assessed: *Moodle, ATutor* and *Sakai*. In 2010, the questioning of higher schools organized by LieDM coordination centre showed that *Moodle* system is used by 18 higher schools, *Sakai* – by 2 higher schools, *ATutor* – by 1 higher school [18]. *Moodle* system is widely used in Lithuanian and world higher schools. *Sakai* system is ranked very high in the world; however in Lithuania, it is little used. The researches of *Sakai* implementation were started in Siauliai University. *ATutor* system is not popular among Lithuanian higher schools; however, it is applied in secondary schools of Lithuania. In 2006, on the ground of this system, the virtual learning environment of Schools' improvement program was created.

The assessment of e-Learning systems included the analysis of the documentation of the systems under consideration on the ground of the research results of various authors [1][2][11]. The application of e-Learning systems for giving joint e-Learning courses was analyzed according to different aspects: the support of multilingual user's interface, the possibilities of giving the course content in several languages, and the functions of the grades conversion into the grading scale of student's institution. The conducted research are summarized in Table 1.

TABLE I. THE EVALUATION OF THE APPLICATION OF E-LEARNING SYSTEMS FOR THE GIVING OF JOINT E-LEARNING COURSES

Criterion	**Moodle 2.4**	**ATutor 2.1**	**Sakai 1.4.3**
Multiple language user interface supports	yes	yes	yes
Navigation between interface languages	yes	no	no
Multiple language content	yes	no	no
Grades conversion function	no	no	no

The conducted analysis showed that e-Learning systems under consideration can work in several language environments. *Moodle* user interface is translated to 112 different foreign languages. *ATutor* supports 71 different foreign languages. *Sakai* supports 20 different foreign languages. Adapting user's interface of the e-Learning system to different languages, menu items and all text variables that can be visible to user are translated. The translation is saved in the separate files that are incorporated in the e-Learning system structure. These files translate the e-Learning system interface, and not the course content.

In *Moodle* system, users can choose the language of interface. User can choose the most appropriate language for him or her from language menu. However, this language choice influences only the interface of e-Learning system.

Multiple language content can be created in *Moodle* system. The multi-language content filter in *Moodle* enables resources to be created in multiple languages.

In e-Learning systems under consideration for the assessment of students' study results it is possible to apply or create different study result grading scales corresponding to the needs of the institution. However, none of the systems under analysis has an integrated grades conversion function which enables the presentation of grades on the student's study results grading scale.

Summarizing the results of the conducted analysis, it is possible to state that the application of e-Learning systems for the giving joint e-Learning courses is not fully implemented. Although the typical feature of *Moodle* system is functional multilingual user's interface and the possibilities of multilingual content creation are implemented, however because of the unresolved questions of the compatibility of study result grading scales, the giving of joint e-Learning courses remains a problematic issue. Thus, e-Learning systems should be improved.

IV. GRADES CONVERSION MODEL

The problem of grades conversion is solved in the works of various authors. In 1997, Haug [12] examined the differences of grading scales used in different countries and emphasized that assessment interpretation is was not more objective than an assessment process itself. The uncertainty of grades conversion is influenced both by the difference of used grading scales and by the different practice of the use of these grading scales. The conversion of grades into ECTS (*European Credit Transfer and Accumulation System*) grading scale received high interest.

Nunes et al. [17] described the method of assessment converting into ECTS grading scale emphasizing separate valid in Portugal converting cases. Warfvinge [20] provided the model of grades conversion into ECTS grading scale.

For conversion of grades in e-Learning systems, two-parameter grades conversion model was created. This model should be applied for conversion of both standard and criteria based grading scale grades. When converting the positive grades from one studies results grading scale to another, the model takes into consideration two parameters, i. e. the distribution of the accumulated control positive grade set data on the grading scale *A* and the distribution of the accumulated control positive grade set data on the grading scale *B*. After labeling the parameters by letters *L*, *K*, the attribution of the equivalent b_j of positive grade a_i on the grading scale *A* to the grading scale *B* can be defined as a two-parameter function (1).

$$a_i = f(L, K, b_j) \quad i = \overline{1,n}, \quad j = \overline{1,m} \qquad (1)$$

The data of grading scales *A* and *B* are written as probability distributions (2). Since in some countries incremental grading scales are used, while decreasing scales are used in others, the marking of the scales is also different and a new variables, i. e., the assessment indexes *i* and *j* are introduced. The assessment indexes number the scale positive grades in the decreasing order and correspond to the characteristic of the distributions.

$$p_{Ai} = P(X = i), \ i = \overline{1,n}, \quad p_{Bj} = P(Y = j), \ j = \overline{1,m} \quad (2)$$

For conversion of assessments from the grading scale *A* to the grading scale *B*, a two-dimensional probability distribution is formed with the values *(i, j)*, $i = \overline{1,\ n}$, $j = \overline{1,\ m}$ (3). The probability p_{ij} of the value *(i, j)* is the probability that the learner's knowledge and abilities, assessed by a grade with index *i* on the grading scale *A*, will be assessed by a grade with the index *j* on the grading scale *B*.

$$p_{ij} = P(X = i, Y = j), \quad i = \overline{1,n}, \ j = \overline{1,m} \qquad (3)$$

The probabilities of the two-dimensional probability distribution are calculated by applying formula (4):

$$\begin{aligned} &p_{ij} = \min\,(p_{Ai} - \sum_{k=0}^{j-1} p_{ik};\ p_{Bj} - \sum_{k=0}^{i-1} p_{kj}), \quad i = \overline{1,n}, \quad j = \overline{1,m} \\ &p_{i0} = 0, \ p_{0j} = 0 \end{aligned} \qquad (4)$$

The positive grade equivalent is attributed on the basis of the formed two-dimensional probability distribution. In case the assessments are not rates, the grade is attributed the most probable equivalent (5).

$$a_i = b_k \text{ , if } \ p_{ik} = \max(p_{i1}, p_{i2}, \ldots, p_{im}), \ i = \overline{1,n} \ (5)$$

In case the probabilities of several grades are equal, the maximum assessment equivalent is used, i. e., the grade with the lowest assessment index. The positive grade a_i on the grading scale *A* corresponds to the positive grade b_k on the grading scale *B*, in case the probability of the distribution value *(i, k)* satisfies the relations, described by equations (6).

$$\begin{aligned} &a_i = b_k \text{ , if } \ p_{ik} = \max(p_{i1}, p_{i2}, \ldots, p_{im}), \\ &\text{if } \ p_{ik} = p_{il}, \text{ then } k < l, \ i = \overline{1,n} \end{aligned} \qquad (6)$$

In case the students are rated, the grades are sorted in the order of decreasing of the assessment value and the corresponding rating is given to each assessment. The number s_{ki} of the assessments, corresponding to the convertible grade a_i is redistributed by applying formula (7). The assessment b_1 corresponds to the s_{i1} of the highest assessments a_i, b_2 – the s_{i2} of the following assessments a_i, etc.

$$\begin{aligned} &s_{ij} = \left[\sum_{k=1}^{j} \frac{p_{ik}}{\sum_{l=1}^{m} p_{il}} * sk_i + 0{,}5 \right] - \sum_{k=0}^{j-1} s_{ik} \ \ i = \overline{1,n}, \ j = \overline{1,m}, \\ &s_{i0} = 0 \end{aligned} \qquad (7)$$

In case the rating of the convertible grade a_i in the assessments set under analysis is *r*, the assessment rating a_i is in the grade group *v*, the number rsk_r of the assessments, corresponding to the rating is redistributed by applying the formula (8). The c^i_{r1} of the highest assessments of rating *r* corresponds to assessment b_1, b_2 corresponds to the c^i_{r2} of the following assessments of rating *r*, etc.

$$\begin{aligned} &c^i_{rj} = \min\,(s_{ij} - \sum_{k=0}^{v-1} c^i_{kj};\ rsk_r - \sum_{k=0}^{j-1} c^i_{rk}), \quad i = \overline{1,n}, \quad j = \overline{1,m} \\ &c^i_{0j} = 0, \ c^i_{r0} = 0 \end{aligned} \qquad (8)$$

The grade a_i of the rating *r* corresponds to grade b_k on the grading scale *B*, in case the relations, described by equation (9) are satisfied.

$$\begin{aligned} &a_i = b_{k,} \text{ if } \ c^i_{rk} = \max\,(c^i_{r1}, c^i_{r2}, \ldots, c^i_{rm}), \quad i = \overline{1,n}, \ k = \overline{1,m} \\ &\text{if } c^i_{rk} = c^i_{rl}, \text{ then } k < l \end{aligned} \qquad (9)$$

E-Learning systems should correspond to the modern tendencies of e-Learning results assessment alteration. In order to implement the idea of learning without walls, to implement virtual student mobility it is necessary to integrate the grades conversion model into e-Learning systems. The results of students' study should be given on the study result grading scale of their country, institution – only then the grades will provide comprehensive reversible information. Thus, continuing the experimental research the grades conversion module was created and integrated into e-Learning system.

V. INTEGRATION OF THE GRADES CONVERSION MODEL INTO E.LEARNING SYSTEM MOODLE

In order to conduct an experimental research, the e-Learning system *Moodle* was chosen. *Moodle* system was chosen by various reasons: software license, reliability, functionality. The results of the conducted research showed that *Moodle* system, in comparison to other open-code e-Learning systems used in Lithuania, is most of all adapted to give joint e-Learning courses.

The grades conversion model was integrated into the *Moodle* system by developing a separate *Moodle* system module. The grades conversion module is a plugin of the *Moodle* system, developed in observance of the rules for and methods of development of the *Moodle* system plugins and corresponding to the *Moodle* plugins technologies. In order to implement the module the PHP (*Hypertext Pre-processor*), HTML (*Hyper text Markup Language*), MySQL and CSS (*Cascading Style Sheets*) technologies were used. The *Moodle* system integrated the grades conversion model by extending gradebook functions and adding a new gradebook report module Fig. 1.

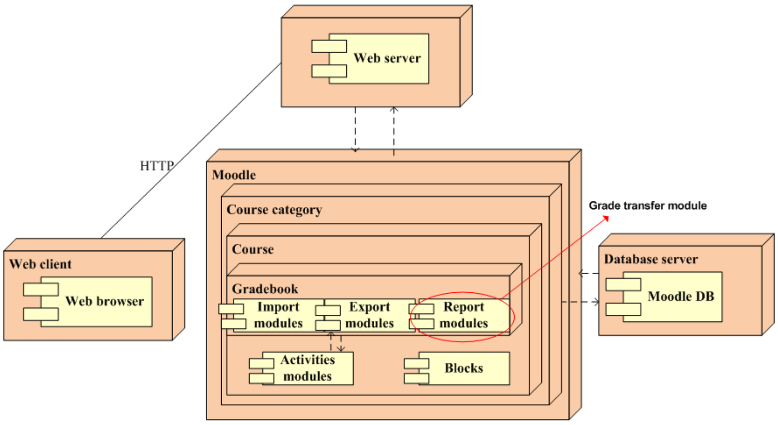


Figure 1. Grades conversion module integration into *Moodle* system components diagram.

The created report module of the gradebook consists of the files intended for module reliability and performance, style, to describe module regulation, to renew the structure of database, to indicate access control.

When integrating the grades conversion module database scheme into the overall structure of the *Moodle* database, the new database tables were developed and the already existing ones were updated. The logical scheme of the database of the grades conversion module is provided in Fig. 2.

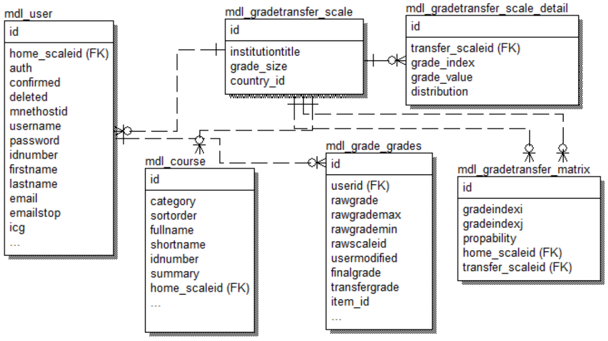


Figure 2. Logical database scheme of grades conversion module.

The report module of the gradebook was created adhering to the general principles of the creation of *Moodle* modules. The main file of grades conversion module is *index.php*. First of all, this file reads configuration parameters, includes necessary libraries, fulfils the check of the conditions of access control, forms the report corresponding to the user's access rights or displays error message.

The reliability and performance of the report module of the gradebook was described creating new, renewing and using standard *Moodle* class methods. Created grades conversion method is executed each time after the recalculation of grades. At the initial stage of grades conversion, necessary grades conversion conditions are checked, i.e. it is established if a grading scale is attached to the course, if there are changeable elements of a course gradebook. If necessary conditions of grades conversion are satisfactory, the grades should be converted for the selected students. During the selection of students, it is checked whether a student got a grading scale which does not match a course grading scale, if a student belongs to a group whose students' grades were recalculated. If the list of students meeting the defined conditions of grades conversion is empty, then grades conversion process is suspended.

After the selection of students, it is checked if necessary grades conversion tables have been formed. If a grades' conversion table corresponding to a course and student grades scales is not made, the method of the formation of this tables is executed. Only those grades of changeable elements of the gradebook are converted which belong to selected students. Making the tables of grades distribution according to ratings, rating process is executed at the group level. The conversion process of the grades of the gradebook is detailed in the grades conversion activity diagram given in Fig. 3.

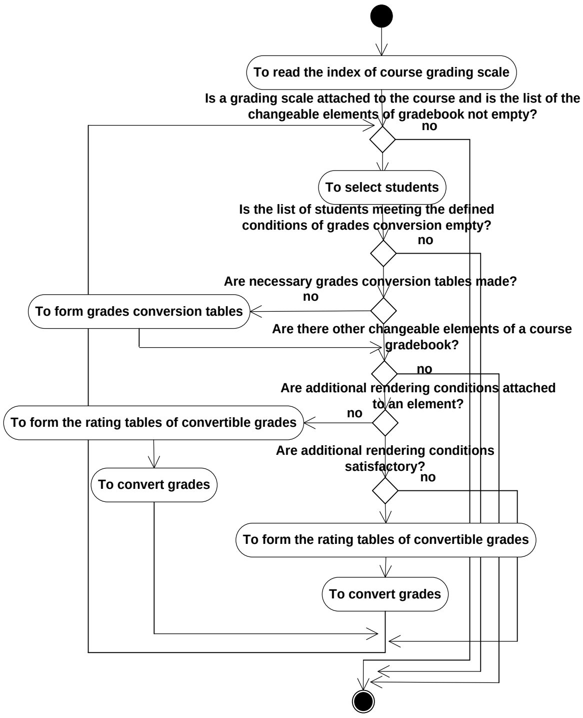
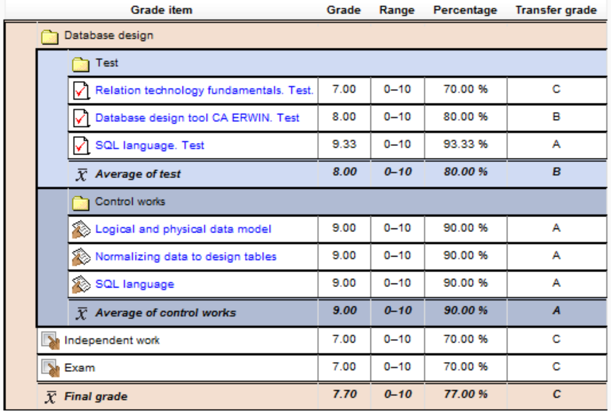

To read the index of course grading scale
Is a grading scale attached to the course and is the list of the changeable elements of gradebook not empty?
no
To select students
Is the list of students meeting the defined conditions of grades conversion empty?
no
Are necessary grades conversion tables made?
no
To form grades conversion tables
Are there other changeable elements of a course gradebook?
no
Are additional rendering conditions attached to an element?
no
To form the rating tables of convertible grades
To convert grades
Are additional rendering conditions satisfactory?
no
To form the rating tables of convertible grades
To convert grades

Figure 3. Activity diagram of gradebook's grades conversion.

When a course teacher defines a course gradebook and establishes the formula of the calculation of the final grade, a student can keep a check on grades changes both on the study results grading scale of the institution giving studies, and on student's study results grading scale. After the student opens the course gradebook, he/she sees the grades on two grading scales. The grades, recalculated to the studies results grading scale of the educational institution, represented by the learner provide comprehensive information to the learner. The student is able to realistically evaluate the achieved results.

A picture given below presents the student's gradebook report. This report shows a student his grades on the grading scale of the institution giving studies (field *Grade*) and also on this student's grading scale (field *Transfer grade*). A picture given below shows gradebook report of student whose institution applies literal A–F study results grading scale (Figure 4.).



Grade item	Grade	Range	Percentage	Transfer grade
Database design				
Test				
Relation technology fundamentals. Test.	7.00	0–10	70.00 %	C
Database design tool CA ERWIN. Test	8.00	0–10	80.00 %	B
SQL language. Test	9.33	0–10	93.33 %	A
$\bar{x}$ ***Average of test***	***8.00***	***0–10***	***80.00 %***	***B***
Control works				
Logical and physical data model	9.00	0–10	90.00 %	A
Normalizing data to design tables	9.00	0–10	90.00 %	A
SQL language	9.00	0–10	90.00 %	A
$\bar{x}$ ***Average of control works***	***9.00***	***0–10***	***90.00 %***	***A***
Independent work	7.00	0–10	70.00 %	C
Exam	7.00	0–10	70.00 %	C
$\bar{x}$ ***Final grade***	***7.70***	***0–10***	***77.00 %***	***C***

Figure 4. Gradebook report of student.

The grades conversion module extends the functionality of the *Moodle* system. It is flexibly integrated into the overall structure of the *Moodle* system. The supplementation of the gradebook with the grades conversion function is relevant when there is a need to give joint e-Learning courses to learners from different countries.

VI. CONCLUSIONS AND FUTURE WORK

Users' interfaces of the open code systems *Moodle*, *ATutor*, *Sakai* used in Lithuania are translated into different foreign languages. The users of *Moodle* system can easily change the language of user's interface; this system has the possibilities of the creation of multilingual content. However, none of the systems under consideration has integrated grades conversion function which is important when giving joint e-Learning courses for students from different countries.

Grades conversion model have been developed. The grades conversion model considers the accumulated grades distribution data on the convertible studies results grading scales and converts the grades by employing the principle of the most probable grades equivalent, taking into consideration the grades rating in the group of the analysed grades.

The grades conversion algorithm is integrated into the *Moodle* system, by expanding the gradebook functions and developing a new *Moodle* system gradebook report module. The scheme of the grades conversion module database is integrated into the common structure of the *Moodle* database, the control of access to the grades conversion model is defined and the functionality of the module is described.

Further we plan to extend grade conversion model including the impact of such factors as language barrier, cultural differences.

REFERENCES

[1] C. C. Aydin and G. Tirkes, "Open source learning management systems in distance learning," The Turkish Online Journal of Educational Technology, vol. 9, pp. 175-185, April 2010.

[2] A. Al-Ajlan and A. Zedan., "Why Moodle," Future Trends of Distributed Computing Systems, pp. 58-64, October 2008.

[3] S. Banks, "Collaboration for inter-cultural e-Learning: A Sino-UK case study" Proceedings of the 23rd Annual Ascilite Conference, The University of Sydney, pp. 71-77, December 2006.

[4] E. Blanchard and C. Frasson, "Making Intelligent Tutoring Systems culturally aware: The use of Hofstede's cultural dimensions," International Conference on Artificial Intelligence, Las Vegas, pp. 644-649, June 2005.

[5] E. Blanchard, R. Razaki, and C. Frasson, "Cross-Cultural Adaptation of e-Learning Contents: a Methodology," International Conference on e-Learning, pp. 112-120, July 2005

[6] S. Chen, Ch. Hsu, and W. Ursuline, "Designing E-learning 2.0 Environment for Cross-cultural Collaborative Learning," Society for Information Technology & Teacher

Education International Conference, pp. 98-105, June 2009.

[7] D. Denev, A. Boichev, and P. Nedelchev. “Multilingual Support of e-Learning Systems: The Bulgarian Contribution,” International Conference on Computer Systems and Technologies - CompSysTech’06, pp. 85-92, June 2006.

[8] A. L. Edmundson, “The Cross-Cultural Dimensions of Globalized E-Learning,” Global Information Technologies: Concepts, Methodologies, Tools, and Applications, pp. 382-393, 2008.

[9] N. Englisch, A. Heller, and W. Hardt, “A Generic Approach for Multilingual Content in Learning Management Systems,” ICERI2013 Proceedings, pp. 312-318 , November 2013

[10] D. Euler, S. Seufert, and F. Moser, “Business models for the sustainable implementation of e-Learning at universities”, Handbook on Information Technologies for Education and Training, pp. 295-315, 2008.

[11] F. Fisler and F. Schneider, “Creating, Handling and Implementing E-Learning Courses Using the Open Source Tools OLAT and eLML at the University of Zurich,” Proceedings of the World Congress on Engineering and Computer Science, pp. 50-57, October 2009.

[12] G. Haug, “Capturing the Message Conveyed by Grades. Interpreting Foreign Grades,” World Education News & Reviews, vol. 10(2), pp. 25-31, 1997.

[13] M. Hillier, “The role of cultural context in multilingual website usability,” Electronic Commerce Research and Applications, vol. 2(1), pp. 2-14, 2003.

[14] J. Lieponiene, “The Research of E-Learning results’ assessment technologies”. Doctoral dissertation, pp. 134, 2012.

[15] M. Mirza and A. Chatterjee, “The Impact of Culture on Personalization of Learning Environments: Some Theoretical Insights,” PLE Conference Proceedings, pp. 56-61, July 2010.

[16] Moodle - Open-source learning platform. [Online]. Available from: https://moodle.org/ [retrieved: February, 2014]

[17] S. Nunes, L. Ribeiro, and G. David, “Supporting the Bologna Process in HE Information Systems,” EUNIS’2005 European University Information Systems, Manchester, pp. 27-33, June 2005.

[18] D. Rutkauskiene, R. V. Musankoviene, and V. Krivickiene, “Common Services Needs for LieDM Network Members,” E-Education: Science, Study & Business, pp. 149-154, November 2010.

[19] The Bologna Process 2020 – The European Higher Education Area in the new decade. [Online]. Available from: http://www.ehea.info/Uploads/(1)/Bologna%20Process%20Implementation%20Report.pdf [retrieved: February, 2014]

[20] P. Warfvinge, “A generic method for distribution and transfer of ECTS and other norm-referenced grades within student cohorts,” European Journal of Engineering Education, vol. 33(4), pp. 453-462, September 2008.

Visual Perception, Speech and Play in the Current Social Tools

On Interactive Technological Device Interfaces

Myzan Binti Noor
MIIT (Malaysian Institute of Information Technology)
University of Kuala Lumpur
Kuala Lumpur, Malaysia
E-mail: myzan@unikl.edu.my

Abstract—The new media interfaces are the visualization that provide text, layout, format, design, user interaction factors and give measurable benefit to the user. That said, the ultimate results are implied through the user's perceptual understanding with the use of the new media devices. This paper emphasizes Vygotsky's work to the present-day use of social tools; i.e. the interactive mobile technology devices. The social tools; interface design - visual communications are children's perception and understandings of learning. The use of the social tools also relate to the constructivism and connectivism over humans instructional instruction with technologies. Therefore, this paper is also interested to examine the link between these theoretical frameworks to Vygotsky's theory of children's speech, play and the zone of proximal development (ZPD) in learning. These fields are important to be examined for children's learning attainments with the technological mediums.

Keywords-new media interface; visualization understanding; social tools; Vygotsky's theory; children's speech; play and the zone of proximal development.

I. INTRODUCTION

This paper examines the relationship of sociocultural activities, such as the use of social tools, to perception, speech, play and social interactions to human learning. The theory of social interactions, speech and the use of social tools was originated by L. S. Vygotsky, (1930) [1]. Back in 1920s of the Soviet Union's era, Vygotsky's work for children's learning included the use of social tools, social interaction, play and educational attainments. His works are mentioned in the learning of theoretical and philosophical related fields in the educational technology literature. The literature environments of children's behaviourism, cognitivism, constructivism, connectivism and instructional instruction and technologies have been presented in today's learning literature. This field has developed tremendously in recent years as part of the social psychology literature related to technology-based learning activity and methodology.

This paper emphasizes Vygotsky's work to the present-day use of social tools; i.e. the interactive mobile technology devices. These social tools; interface design - visual communications are children's perception and understandings of learning. The use of these social tools also relate to the constructivism and connectivism over humans instructional instruction with technologies.

Therefore, this paper is also interested to examine the link between these theoretical frameworks to Vygotsky's theory of children's speech, play and the Zone of Proximal Development (ZPD) in learning. These fields are important to be examined for children's learning attainments with the technological mediums.

II. THE STATE OF ART

The new media interfaces are the user-generated content that is shared. The research on the new media is focused now on collaboration, connections, emotion and communication [2]. Hence, the focus is on understandings people's needs, i.e., learners' preferences, likings, techniques and the different requirements for learning and support. Moreover, with the use of tablet PCs in the classrooms, sharing of information and so on, children and teachers collaborate in learning via wireless networking [3]. This paper focuses on these preferences which are still lacking. The importance of the new media technology devices is undeniable to the learners. The urges for this paper is timely in upholding the notions of social interactions theory of Vygotsky that learning is a concept of guidance and interaction, the ZPD concepts, and conversation and play.

III. THE RESEARCH PROBLEM

The new media technological devices are the digital-age social tools. The medium connects a human user to electronic information of any kind. The new media interfaces are the visualization understandings that provide text, layout, format, design, user interactions factors that yield measurable benefit to the user. That said, the ultimate

ISBN: 978-1-61208-342-1

results are implied through the user's perceptual understandings with the use of the new media devices. Therefore, it is important that this paper stresses the instruction of the interface design with the theory of perception and speech when children learn. The instructional requirements are the HCI (Human–Computer Interactions) factors that contribute to the human cognitive process involved and cognitive limitations. Besides, human needs supply theories; interface designs, modelling tools, guidance and methods that can lead to the design of better interactive products [4]. Diana Laurillard [5], for example, highlights the importance of the academic goal rather than just imparting knowledge. A teacher's skill of providing knowledge is crucial, that is his or her higher aim for learners to achieve the acquired knowledge.

The results show that learners are lacking on the subject of specific experience, access needs, motives for learning, expectations, prior experience and preferred approach, and many more [6]. So far, research has been done so much on the task performances, statistical software, etc. But, the aspects on the accessibility and inclusion for learners-centred approach to support individual needs are lacking.

This research stresses the considerations to learners are underpinned over the interfaces usability, messages and interpretations that contribute to children's understandings with the teachers help with speech, teaching guidance and the use of the internet tools. The technological tools are such as the interactive white boards, the wikis, the blogs and the RSS feeds. These technological tools are the medium that this generation has been upholding contentedly for their knowledge. They acquire the new media like the 'bread and butter' in gaining knowledge and other interests in almost everything they do.

IV. THE SIGNIFICANCE OF THE STUDY

Learners' perception, visual communications and new media

The social contexts are important in determining how technology might influence teaching and learning. This statement is argued by Roy Pea, director of the SRI Center for Technology in Learning in Menlo Park, California. The technology is crucial to be used in understanding learning must consider on the strategies that relate to the social contexts. Special attention should be paid to the teaching strategies used both "in" the software and "around it" in the classroom, and to the classroom environment itself.

It is a recurrent finding that the effects of the best software, visual message and narrative can be neutralized through improper design, or use, in learning outcomes.

Learners and instructors' speech

In mentioning the children's learning achievements, Vygotsky believes human speech is the most important sign-using behaviour in children's development. He claims that children free themselves through speech from many of the immediate difficulties that they are facing. With speech, children are able to prepare, plan, order and control their own behaviour as well as others for future activities (Cole, et al., 1978) [7]. In an experiment, Vygotsky recorded that a child achieved their goals just by using speech. A child's speech and actions go on simultaneously; thus, he or she provides commentary after speaking about what they are doing. The commentary describes how a child engages in a number of initial acts, as well as mediated methods such as asking questions to the people standing nearby them. At the same time, the child solves his or her problems as their speech reacts to their complex psychological functions of perception, attention and social interaction. Hence, this paper relates the children's social interaction through the 'external stimuli' of the technological devices. Children use the devices as their source of operating learning activity. The child speech is an activity of their external activity that develops their inner organization of thought from the new media/technological devices. The speech is the inner organization of thought which stimulates, mediates and regulates the child's daily activity, including learning. In turn, those thoughts mediate the meaningful signs of their speech and actions. Speech and actions, Vygotsky claims, mediate the child's thinking to a much higher level of intellectual. This means, the child is able to develop, engage, respond and produce intellectually productive learning, work in the classroom, house and society. He asserts that the greater the child's action, the more they rely on speech. Vygotsky wrote:

> *"They are characterized by a new integration and co-relation of their parts. The whole and its parts develop parallel to each other and together. We shall call the first structures elementary; they are psychological wholes, conditioned chiefly by biological determinants. The latter structures which emerge in the process of cultural development are called higher structures... The initial stage is followed by that first structure's destruction, reconstruction, and transition to the structures of the higher type. Unlike the direct, reactive processes, these latter structures are constructed on the basis of the use of signs and tools; these new formations unite both the direct and indirect means of adaptation"* [8]. (Vygotsky, 1930)

Thus, how does the speech relate to children's perception and visual communications then contributes to their learning development? Vygotsky argued that the higher psychological functions of human development are such as perception, attention, speech; sensory-motor operations and memory will eventually form the unity of children's goals and adaptation. For example, his student, A. R. Luria stated that the unitary functions of these components are formed during each individual's development and are dependent upon the social experiences of interactions in the child's environment and culture. Then, the functional systems of an adult are shaped essentially by his or her prior experiences as a child, as well as the social aspects (Cole, et al., 1978) [9]. The basic functions are integrated into new functional learning systems. At the same time, the child's higher psychological functions are not being usurped by the basic processes. They represent a new psychological system in the child that leads to their intellectual understandings.

The role of Play in Children's Development

Vygotsky claims that play advances the development of a child. Play is important to children – they satisfy certain needs in play by using their imaginations and acting out their desires immediately. Vygotsky argued that children perform as an adult character during play. They act out the activities of their culture and rehearse any future roles and values that they admire. Vygotsky explained [10]:

> *"The child sees one thing but acts differently in relation to what he sees. Thus, a condition is reached in which the child begins to act independently of what he sees".* (Cole, et al., 1978, p. 97)

Vygotsky asserts here that everyday situations in a child's behaviour are the opposite of his behaviour in play. He claims that during play, a child's actions are subordinated to meaning, but in real life, action dominates meaning. In play, a child always behaves beyond his or her age, above himself or herself. All children's developmental tendencies are condensed and contained in play. So, in play, children's development can be compared to instruction. Vygotsky further claims that play provide a much wider background for changes in needs and consciousness. Children's action in play is their imaginative sphere of the world where they create the voluntary intentions, i.e. they act as a doctor, a teacher or someone they looked-up to. The creation of voluntary intentions such as those mentioned, could affect the formation of their future real-life plans and desirable motives. Desirable motives are the imagination of the roles that they play. For example, children play with the doctor's first aid kit, one of them acts as a patient, and the doctor is looking after the patient, and so on. Vygotsky claims that play, moves children forward significantly and is the highest level of their preschool development. Therefore, play is considered a leading activity that determines the child's development (Cole, et al., 1978) [11].

Vygotsky argued that human activity transforms both nature and society. During preschool and school years, the conceptual abilities of children are stretched through play and the use of their imagination. The play-development relationship is not like an instruction - development relationship. Play gives a broader background for changes in needs and consciousness to children [12]. By the early age of human development, they have already experienced the tension between desires that can be fulfilled only in the future and is something that demands immediate enjoyment. Through play this contradiction is explored and temporarily resolved. Here, Vygotsky places imagination as representing a specifically human form of conscious activity. It arises from the action in play.

V. THE RESEARCH QUESTION

Our research question is focused on: *1. How do the child's perception, speech, play and the use of social tools relate to their visualization understandings? 2. What type of consideration the adults/teachers should look into in helping them?*

This research deals with the approaches in helping children visualization with new media by the adults; i.e. teachers, parents, peers can help children's learning. For example, Vygotsky emphasised the role of learning in facilitating and supporting people in society. In the societal context, children's learning and behaviour development require interaction and support by their guardian, teacher, more capable peers or parents. Interaction, engagement and participation are the major factors that are critical for children's learning process. Children learn from teachers, more capable peers or parents. Vygotsky included the specification of the societal context in which the child's behaviour developed, such as children's action or reaction through their involvement with society that develops the quality of humans and relationships.

VI. THE RESEARCH METHODOLOGY

Research has shown that the human-computer interactions field has not been led to a specific textbook [13]. Thus, teaching and learning instructional methodology that put emphasis on the HCI and ZPD concepts in the classrooms with the technological devices has yet been pioneered. This paper proposes a few research questions that help underpin the research contribution to children's learning. Preece, Rogers and Sharp [14], claim that attention, perception and recognition, memory, reading, speaking and listening, problem-solving, planning, reasoning and decision-making learning are important. These are the core cognitive aspects. Moreover, learners' preferences are important to be recognised and examined. Dagger, Wade and Conlan [15] discussed about learners preferences such as different concerns, likings, techniques and different needs for learning and support.

For example, this paper highlights the children's developmental process of perception, speech, play and the use of social tools with the technological mediums relate to the children's perceptual understandings. A few case studies will be examined based on the learning activity programs that children and teacher have developed in the classrooms in the European countries. A detail of the case studies will be presented to show the documentation of the learning activity concerning visual communications field studies. This includes the principle and elements of the arts, i.e., colours, lines, space, contrast, etc. The methodology includes observation, test, questionnaire and interview sessions with school children. The targeted children are between 10-15 years old.

This research proposes the adults, i.e. parents, teachers and peers to help children's learning. Children's cognitive development process, Vygotsky argued, is a process of 'telescopes changes'. Telescope changes are observing the process of children's actions and reactions (Cole, et al., 1978). Therefore, a special program for children's learning approach will be set to examine children's learning styles in the classrooms. At home, the task is for the parents to monitor and record. Thus, the results of the teacher/parent monitoring will be documented and examined to formulate the outcomes. Vygotsky's concept of ZPD emphasises the various sociocultural structures and their impact on the interactions between individuals, artefacts, technology and environment. Vygotsky asserts that teaching is stimulated by insightful development and subsequent learning. He argues that teaching means providing advancement to the learner, socially elaborated human knowledge and cognitive development. Cognitive development is something for which learners must put their own reflective and internal strategies to work. Vygotsky named the "actual developmental levels" which characterise mental development retrospectively. Simply put, if a child can do such-and –such independently it means the functions have already matured in the child. And, the ZPD is the assistance provided to the child when he or she cannot do such-and-such independently.

VII. THE CONTRIBUTION TO KNOWLEDGE

This research is about the perception, speech, and the use of social tools in children's development. Vygotsky's research shows that there is a link of mediated activity in between the goal and the reaction in children. He claims that children's higher psychological processes of memory require children's reaction in order to produce something. Vygotsky considers perception, attention and speech as the children's reaction. Thus, these functions help promote their learning. With these functions, children produce action reaching their goals. Vygotsky argued that through such stimuli, a child would be able to see the immediate situation and react upon it. Vygotsky describes it as 'active human intervention' (Cole, et al., 1978) [16]. Vygotsky claims that these supporting stimuli are for children a means of active adaptation. The supporting stimuli are highly diverse and include the tools of the child's culture, the language of those who relate to the child and the ingenious means produced by the child himself, including the use of his own body (Cole, et al., 1978) [12].

VIII. CONCLUSION

As such, high levels of engagement can in turn affect the cognitive distribution of children's perception, understanding and confidence for educational achievements. Their attention, inquisitiveness and reflection are developed in this context. The arguments have sustained many of the examples made by scholars of social interaction - using intellectual development as the factors of stimuli in children's cognitive growth. Vygotsky uses the example of play by poor children who do not have access to manufactured toys but whom, nevertheless, are able to play house, train, and so on with whatever resources are available to them. Theoretical explorations of these activities in a developmental context are a recurrent theme of this thesis. Similarly, cognitive development in children's perception, understanding and confidence for learning should be looked into.

ACKNOWLEDGMENT

I wish to acknowledge my appreciation to the dean of MIIT, Prof. Dr. Roslan Ismail, the University of Kuala Lumpur President Prof. Dato' Dr. Mazliham Mohd Su'ud and the staff of UniKL for their help and support.

REFERENCES

[1] M. Cole, V. John-Steiner, S. Scribner, and E. Souberman, "L.S. Vygotsky, Mind in Society: the development of higher psychological processes", Cambridge MA: Harvard University Press, London, England, 1978.

[2] L. Jonathan, J. Feng, and H. Hochheiser, "Research Methods in Human – Computer Interaction", John Wiley & Sons Ltd, 2010.

[3] D. Dagger, V. Wade, and O. Conlan, "Personalization for all: making adaptive course composition easy", Educational Technology and Society: Special Issue on Authoring of Adaptive Hypermedia, 8 (3), pp. 9-25, 2005

[4] J. Preece, Y. Rogers, and H. Sharp, "Interaction Design: Beyond Human Computer Interaction", John Wiley & Sons, 2002.

[5] D. Laurillard, "Rethinking University Teaching – A framework for the effective use of learning technologies", 2nd edition, London: Routledge/Falmer, 2008.

[6] H. Beetham and R. Sharpe "Rethinking Pedagogic for a Digital Age, Designing and Delivering e-learning, an Introduction to Rethinking Pedagogy", Routledge Taylor & Francis Group, London and New York, 2007.

[7] M. Cole, V. John-Steiner, S. Scribner and E. Souberman, "L.S. Vygotsky, Mind in Society: the development of higher psychological processes", Cambridge MA: Harvard University Press, London, England, 1978.

[8] Ibid, pp. 102.

[9] Ibid, pp. 19-30.

[10] Ibid, pp. 97.

[11] Ibid, pp. 92-104.

[12] Ibid, pp. 38-51

[13] L. Jonathan, J. Feng, and H. Hochheiser, “Research Methods in Human – Computer Interaction”, John Wiley & Sons Ltd, 2010.

[14] J. Preece, Y. Rogers, H. Sharp, “Interaction Design: Beyond Human Computer Interaction”, John Wiley & Sons, 2002.

[15] D. Dagger, V. Wade and O. Conlan, “Personalization for all: making adaptive course composition easy”, Educational Technology and Society: Special Issue on Authoring of Adaptive Hypermedia, 8 (3), pp. 9-25, 2005

[16] This passage is from the unedited translation of “Tool and Symbol”, 1978. “L.S. Vygotsky, Mind in Society: the development of higher psychological processes”, edited by M. Cole, V. John-Steiner, S. Scribner and E. Souberman, Cambridge MA: Harvard University Press; London, England, 1978.

Intelligent Multimedia Mind Maps to Support Media Pre-Production

Erik Mannens, Ruben Verborgh, Rik Van de Walle
ELIS – Multimedia Lab
iMinds – Ghent University
Ghent, Belgium
{erik.mannens, ruben.verborgh, rik.vandewalle}@ugent.be

Simon Debacq, Maarten Verwaest
Limecraft
Ghent, Belgium
{simon.debacq, maarten.verwaest}@limecraft.com

***Abstract*—To date, there are almost no tools that support the elaboration and research of project ideas in media pre-production. The typical tools that are being used are merely a browser and a simple text editor. Therefore, it is our goal to improve this pre-production process by structuring the multimedia and accompanying annotations found by the creator, by providing functionality that makes it easier to find appropriate multimedia in a more efficient way, and by providing the possibility to work together. To achieve these goals, intelligent multimedia mind maps are introduced. These mind maps offer the possibility to structure your multimedia information and accompanying annotations by creating relations between the multimedia. By automatic connecting to external sources, the user can rapidly search different information sources without visiting them one by one. Furthermore, the content that is added to the mind map is analyzed and enriched; these enrichments are then used to give the user extra recommendations based on the content of the current mind map. Subsequently, an architecture for these needs has been designed and implemented as an architectural concept. Finally, this architectural concept is evaluated positively by several people that are active in the media production industry.**

Keywords - media pre-production; mind maps; information search; semantic web.

I. Introduction

In professional media pre-production [1], there is little support to elaborate on an idea. Usually, one only uses a browser and a text editor as tools, i.e., one to search for information and one to gather the information [2], but most of the idea is in their brain and not on virtual and/or collaborative "paper".

The main problem with the current research method is that there is almost no logical structure, as most of the structure is in the brain of the creators. In a co-production, this problem becomes even bigger, because the idea is spread over multiple brains. A result of the lack of structure is that it is hard to reuse information in future productions, because existing documents –if they still exist at all– are hard to comprehend, since it is a linear list of non-related pieces of information, and thus, it is hard to find the necessary information, certainly for people that did not perform the research in the first place. Another problem is the fact that information is widely spread, and thus, creators have to use several search engines within different distributed information bases.

We counter these problems by introducing the notion of *intelligent multimedia mind maps* [3]. First of all, this multimedia mind map structures the found multimedia. This structure, as such, has little short-term impact, but has a big long-term influence, since it is easier to reuse older work as re-finding sources is self-evident by reusing the accompanying annotations. Secondly, the information from the multimedia mind map can be used to suggest new relevant information; hence from now on our implemented automatic recommendations makes us talk about "intelligent" multimedia mind maps.

Section 2 describes the concept of mind mapping, whereas Section 3 explains why current search services will be reused. Furthermore, Section 4 elaborates on the Architectural Concept, and afterwards, Section 5 evaluates our solution. Finally, Section 6 draws conclusions and looks at future extensions.

II. Mind Mapping

A mind map [4] is a tree structure where you start with a main topic, preferably the center of your mind map. Subsequently, you associate subtopics with the main topic; thereafter you do the same with the subtopics. To have a view on what's available we searched for existing mind map software. There are plenty, but most of them only support text. However, there are two that have more functionality. Mind42 [5] furthermore supports images and collaboration, but lacks support for videos and connection with external sources. On the other hand, Visual Understanding Environment (VUE) [6] supports images and limited connection with not-modifiable external sources, but lacks support for videos, and collaboration. Our envisioned intelligent multimedia mind maps do fully support text, images, audio & video, external links via Linked Open Data (LOD), and collaborative user management.

III. Search Service

There are many multimedia sources and most of them have a search service, which is optimized for their own needs. Both local and generic services are apparent, i.e., an example of such a local service is MediaLoep [7], which provides a service to search the VRT [8] video archive. The most well known generic multimedia search engine is Google's YouTube [9]. Therefore, it is not our goal to provide our own search service but to integrate existing services by using and incorporating their Application Programming Interfaces (API).

IV. Architectural Concept

Before creating an architectural concept, a generically extensible architecture was designed according to the generic requirements, resulting from the analysis of the problems in current methods of topic research in media pre-production, and conforming the Attribute-Driven Design (ADD) [10] principles. This resulted in a 3-tiered layer architecture, as shown in Figure 1. The top-layer communicates with the client and forwards commands to the second layer, i.e., the model layer, which –among other things– comprises the state of the mind map and uses the services from the bottom layer, the service layer. There are three main services in the bottom layer, i.e., storage, enrichment, and recommendation. The storage service is a module that is used to communicate with a database. In this case, a graph database was chosen, because of better performance on traversing related data, and greater flexibility [11].

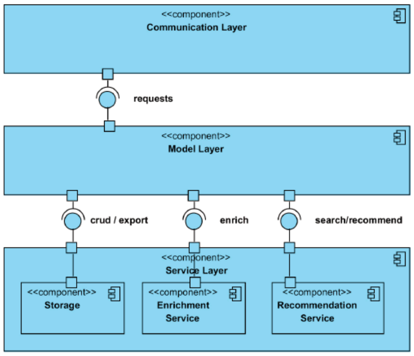


Figure 1. Generic Architecture of the Architectural Concept.

The enrichment service is decomposed according to a composite pattern [12]. This way it is easy to add and remove enrichers, which are modules that analyze the content and return important features. For the architectural concept Named Entity Recognition (NER) was performed using Apache Stanbol [13], which was chosen over DBpediaSpotlight [14] as such, because the first includes the DBpediaSpotlight service and gives more possibilities with modifiability in mind, e.g., you can incorporate your own thesaurus and/or ontology. Mind you, in the current architectural concept there is only a text enricher, but this can easily be extended to enrichers for other multimedia formats.

The recommendation service is similar to the enrichment service and is also decomposed into a composite pattern. This makes it easy to always add extra modules that connect with external sources with their accompanying search functionality. To be able to show this functionality in the architectural concept three modules, i.e., a connection with Flickr [15], Wikipedia [16] & its LOD counterpart DBPedia [17], and YouTube, were implemented. These three initial sources were chosen because they offer three different multimedia formats and they have a publicly available and well-documented API. In Figure 2, you can see an overview of the resulting architectural concept application.

V. Evaluation

Firstly, our predefined goals –especially the goal to improve the structuring and reuse of the content– were confirmed by the people of both Limecraft [18] and Taglicht Media [19]. They both have the need for more structure in their research, because it is hard to find content again, certainly when they share their research with others.

Structuring content according to a mind map structure is a beginning, but there was need for more functionality, including named relationships and cross-linking. The integration of external LOD sources was a great addition, because the creator can simply add a piece of content to the mind map, but the current implementation has it flaws. The user always has to add the complete piece of content, but is usually only interested in a quote or a small piece of the content. Another remark is that it should not only be possible to search external sources, but also the content that's in some other mind maps. This would be useful in very large mind maps, certainly when you don't know how the content was structured in the first place.

Also the recommendation functionality is a nice feature to have, but there are some limitations as well. It cannot replace the human brain and therefore, it is only suited to define high-level concepts, which is useful to rapidly divide research work. It cannot, at least for now, give the user a full background of his developing topic.

VI. Conclusion and Future Work

The provided solution for the predefined goals show potential, certainly the structural functionality as this is a big problem right now. Because reuse of research data saves a lot of time, it has a big impact on the value chain of future productions. However, before our tool can be used in professional media production, there's a need to improve the application taking the feedback of the previous section into account, i.e., the incorporation of media fragments and the search between implemented intelligent mind maps.

Next to the improvements to the current application there's also the possibility to add some extra functionality. The mind map could be more intelligent by adding the possibility that when you add a piece of content you get suggestions how this piece of content is related to the content in the current mind map. For example, when you have two nodes as pictured in Figure 2, one about "Caesar" and one about "Pompey", and you add a node about "Julia" you get the suggestion to add it to the node about "Caesar" with relation 'daughter' and/or to the node about "Pompey" with the relation 'wife'.

Other extensions are to provide export functionality for the multimedia mind maps to support interoperability or to add the functionality to create a scenario. Merging the two applications –scenario creation and multimedia mind map– in one application makes it easy to switch between creating the scenario and expanding the mind map, which is useful because these processes happen in an iterative and parallel way anyway [20]. Another advantage would be to drag and drop parts of the mind map, e.g., quotes, into the scenario.

ACKNOWLEDGMENT

The research activities that have been described in this paper were funded by Ghent University and iMinds (Interdisciplinary Institute for Technology) a research institute founded by the Flemish Government.

REFERENCES

[1] L. Hardman, Z. Obrenovic, F. Nack, B. Kerhervé, and K. Piersol, “Canonical processes of semantically annotated media production”, ACM Multimedia Systems Journal, vol. 14, no. 6, September 2008, pp. 327–340.

[2] D. Van Rijsselbergen, B. Van De Keer, and R. Van de Walle, “The canonical expression of the drama product manufacturing process”, ACM Multimedia Systems Journal, vol. 14, no. 6, September 2008, pp. 395-403.

[3] M. Davies, “Concept mapping, mind mapping and argument mapping: what are the differences and do they matter?”, Journal of Higher Education, vol. 62, no. 3, September 2011, pp. 279-301.

[4] M. J. Eppler, “A comparison between concept maps, mind maps, conceptual diagrams, and visual metaphors as complementary tools for knowledge construction and sharing”, Information Visualization, vol. 5, no. 3, September 2006, pp. 202–210.

[5] Mind42 [Online]. Available from: http://mind42.com/ [retrieved: April, 2014].

[6] Visual Understanding Environment [Online]. Available from: http://vue.tufts.edu/ [retrieved: April, 2014].

[7] P. Debevere, D. Van Deursen, E. Mannens, R. Van de Walle, K. Braeckman, and R. De Sutter, “MediaLoep: Optimizing search in a broadcaster archive”, Proceedings of the 12th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2011), April 2011, pp. 1-2.

[8] VRT - Vlaamse Radio & Televisie - Belgian’s regional broadcaster [Online]. Available from: http://www.vrt.be/ [retrieved: April, 2014].

[9] YouTube [Online]. Available from: http://www.youtube.com/ [retrieved: April, 2014].

[10] L. Bass, P. Clements, and R. Kazman, “Software Architecture in Practice”, Addison-Wesley, third edition, October 2012.

[11] C. Vicknair, M. Macias, Z. Zhao, X. Nan, Y. Chen, and D. Wilkins, “A comparison of a graph database and a relational database: a data provenance perspective”, Proceedings of the 48th annual Southeast regional conference, April 2010, pp. 42.

[12] E. Gamma, R. Helm, R. Johnson, and J. Vlissides, “Design Patterns: Elements of Reusable Object-Oriented Software”, Addison-Wesley, November 1994.

[13] Apache Stanbol [Online]. Available from: http://stanbol.apache.org/ [retrieved: April, 2014].

[14] P. N. Mendes, M. Jakob, A. Garcia-Silva, and C. Bizer, “Dbpedia spotlight: shedding light on the web of documents”, Proceedings of the 7th International Conference on Semantic Systems, September 2011, pp. 1–8.

[15] Flickr API [Online]. Available from: http://www.flickr.com/services/api/ [retrieved: April, 2014].

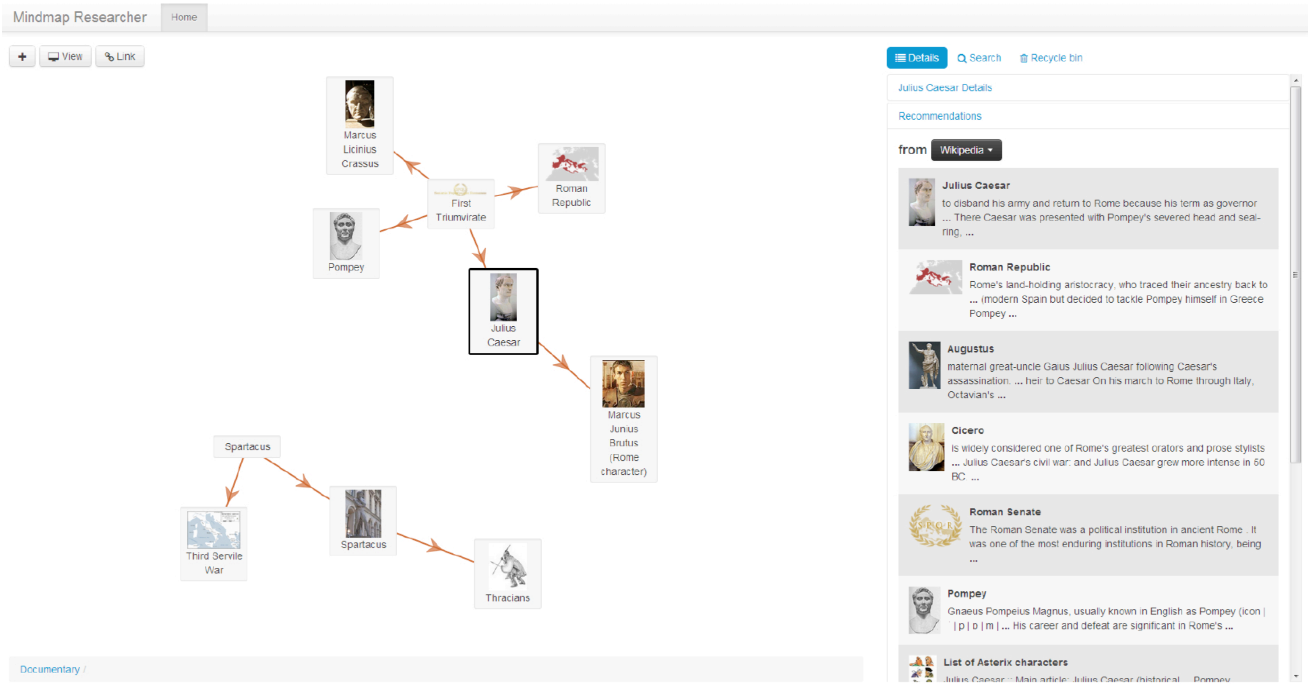
[16] Wikipedia API [Online]. Available from: http://www.mediawiki.org/wiki/API [retrieved: April, 2014].

[17] DBPedia API [Online]. Available from: http://wiki.dbpedia.org/lookup/ [retrieved: April, 2014].

[18] Limecraft [Online]. Available from: http://www.limecraft.com/ [retrieved: April, 2014].

[19] Taglicht Media [Online]. Available from: http://www.taglichtmedia.de/en/ [retrieved: April, 2014].

[20] A. Rosenthal, “Writing, Directing and Producing Documentary Films and Videos”, Southern Illinois University Press, June 2007.



Julius Caesar
to disband his army and return to Rome because his term as governor ... There Caesar was presented with Pompey's severed head and seal-ring, ...

Roman Republic
Rome's land-holding aristocracy, who traced their ancestry back to ... (modern Spain but decided to tackle Pompey himself in Greece Pompey ...

Augustus
maternal great-uncle Gaius Julius Caesar following Caesar's assassination. ... heir to Caesar On his march to Rome through Italy, Octavian's ...

Cicero
is widely considered one of Rome's greatest orators and prose stylists ... Julius Caesar's civil war: and Julius Caesar grew more intense in 50 BC. ...

Roman Senate
The Roman Senate was a political institution in ancient Rome . It was one of the most enduring institutions in Roman history, being ...

Pompey
Gnaeus Pompeius Magnus, usually known in English as Pompey (icon | ˈ | p | ɒ | m | ... His career and defeat are significant in Rome's ...

List of Asterix characters
Julius Caesar :: Main article: Julius Caesar (historical ... Pompey

Figure 2. Architectural Concept Multimedia Mind Map Application.

Refined Ontology Matching Methods for Special Data Integration

Dénes Paczolay, András Bánhalmi, Ádám Zoltán Végh, Gábor Antal, Vilmos Bilicki
Department of Software Engineering
University of Szeged
Szeged, Hungary
{pdenes,banhalmi,azvegh,antalg,bilickiv}@inf.u-szeged.hu

***Abstract*— Ontology matching is an important area of research, since it has many applications like semantic webs, information extraction, data mining and reasoning. In most cases, the matching is done between thesauri of hierarchical concepts made for similar domains by different groups. A methodically similar, but technically different task is when ontology matching is used for system integration applied to generated ontologies. For data integration, we created a framework that generates ontologies got from a database schema or from the source code itself, these being called local and global ontologies. To complete the data integration process, only ontology matching has to be done semi-automatically; all the other tasks are carried out automatically. However, the generated ontologies have some special features, namely mixed languages in the name, special abbreviations, and special structures in the names generated. For ontologies like these, the common matching methods which have the best performance on average when these are applied to other tasks, perform much worse in the case of integration. In order to improve the accuracy, we propose novel similarity measuring and ontology matching methods.**

Keywords- semantic knowledge representation; local ontology generation; ontology matching; ontology alignment; ontology; Java to ontology; similarity measure

I. INTRODUCTION

In the past few decades, many ontologies have been created to describe semantically different specific domains. The ontologies created for a similar domain by different groups contain the same or similar concepts, and connections between them in most cases. The fact that there are many similar ontologies implied the need for querying all the ontologies at the same time with the same expression. For this, many research groups proposed to create a mediator schema, a global ontology, which integrates the necessary concepts got from all the so-called local ontologies. In addition, for the purpose of creating a broader semantic search, not only ontologies and RDF (resource description framework) stores, but information from conventional RDBM (relational database management) systems and (nowadays) NoSQL (no structured query language) data stores are also involved. The aim of system integration is similar in some sense: applications with a similar functionality should be able to exchange their data with each other. This task can be resolved by hardcoding the data transformation and calling the services of each other, but this solution is time-consuming, the productivity of development is low, and the reengineering is difficult. For this reason, the concept of semantic integration can be used to integrate the data of the systems with a similar functionality. Some methods and techniques have already been proposed for data integration using ontologies. These solutions may contain one mediator schema only, which integrates all the databases virtually, or contain more ontologies from which one mirrors the concept of a database, the other ontology being responsible for the semantic integration. These are called local and global ontologies. In our system integration framework, the local ontologies are generated using semantic information collected from the database schema itself. The global ontology may also be a generated one; moreover, in one of our use cases it is generated from the source code of the integrating central application itself. The principal problem, however, is that a more robust ontology matching method is needed among the concepts of these specially generated ontologies for the purpose of more efficient and quick data integration development. For the ontology matching problem, many methods have been proposed and implemented, but these methods are too general for handling special issues concerning the rules of the ontology generation and frequent naming cases. To create a more robust ontology matching process, we propose new similarity measures for generated ontologies, and carry out some experiments that apply them in real-world tasks to evaluate their precision.

In the sections below, a detailed overview of related work is provided. Then, we introduce and discuss the ontology generation methods, and many type of problems are examined that are related to ontology matching. After, some similarity methods are proposed, which can improve the performance of ontology matching for generated ontologies. In our experiments, real-world use cases are introduced, and the results of the proposed methods are then compared with those got using the known matching techniques. Lastly, we discuss our experimental results, and make some suggestions for future study.

II. RELATED WORK

The chief goal of ontology matching is to collect all the knowledge or information related to a common domain. Because many ontologies were created for a similar domain, an integration method was needed to collect interesting data from them. Integration can be achieved by finding relations among the concepts [1][2]. What these solutions have in common [3]–[6] is that in general, an iterative matching

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-342-1

process is performed: after each iteration the result of a matching process can be the input for the next matching process. Also, the results of the matching processes are aggregated. This mechanism can be arranged hierarchically to implement a more complex ontology matching. The basic concept similarity measure used by ontology matching methods is to compare the strings describing the concepts. For this task string matching methods are used [5][7][8], perhaps completed with some additional, but sometimes important complex natural language processing (NLP) methods (translation dictionaries, wordnets, Tf/Idf (term frequency / inverse document frequency) categorization, etc.) [3][9][10]. Besides language processing methods and string matching, many graph-structure based investigations have been proposed to aid the matching process [11][12].

In addition to the above-mentioned ontology merging tasks, other applications also apply the so-called ontologies to describe semantic information about specified objects, and in this way achieve an integrated result. In this area, the integration of Web services should be mentioned [13], and data (or system) integration is also an important application domain [14][15]. In the latter (which is the focus of this paper), the ontologies for a mediation of local and global information are widely used, but we do not have any information about ontology matching methods, which have been developed and focus on generated ontologies, considering, for example, the generation rules, and the naming variations in practice. The widely-used heterogeneous benchmark, which is used for Ontology Alignment Evaluation Initiative (OAEI) [16], does not contain ontologies of this kind [2].

III. USE CASES

An industrial partner suggested that they would like to change their POS printing system to a new one, because their original system had become overly complicated, so adding new clients to the system was hard to implement. To support easy system integration, an ontology-based integration framework was developed. By using different open-source tools with some important modifications added by us, and implementing new modules (e.g., for query rewriting), a well-functioning tool chain was developed. Using this technique for integration, only one task needed human assistance, namely to find the relations among the concepts of the local and the global ontologies. The local ontology was always generated corresponding to the schema of the database. The global ontology was constructed by humans in the “POS Printing” use case, or it was generated from Java source code for the “Event Integration” use case.

A. *POS Printing Data Integration*

The starting point in this application is a PostgreSQL relational database containing about 70 tables to describe products, department stores, promotions, printing templates, etc. The endpoint is a global ontology assembled by two people, which covers the topics of products, promotions and printing. This global ontology was created in two languages, namely English and Hungarian. The local ontology vocabulary was generated from the schema of a relational database using the D2RQ generator [17] with some minor modifications which are described here [19]. Fig. 1 illustrates a little snippet of the global ontology created for this use case.

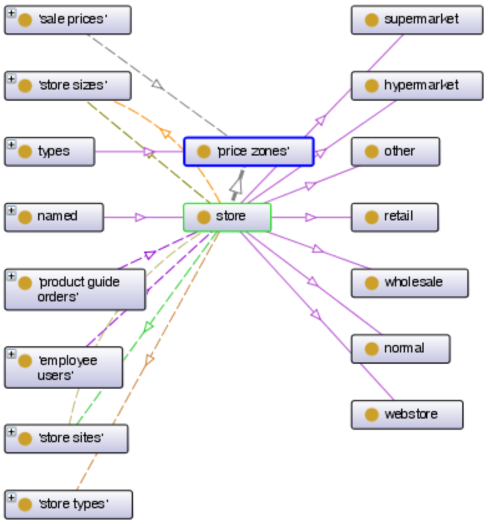


Figure 1. A snippet from the global ontology created for the POS printing use case.

B. *Calendars and events*

In another use case for system integration, we chose calendar integration. This means that it should be possible to collect event-like data got from various applications for the Google Calendar. For this, we searched scheduling, time/event tables, and booking applications with database support on websites such as SourceForge, Google code and Hot Scripts. In the end, we collected 16 applications with their sources. The selected programs were very diverse; some of them used 2-3 tables only, and others had over 15 tables. From the databases, the local ontologies were generated using a modified D2RQ generator [19].

The global ontology for describing the concepts in the Google Calendar was generated from the Java source code of Google Calendar Client API. For this purpose, a Java2RDF application was developed. This application uses QDox [18] to explore class, interface and method definitions.

IV. METHODS

Here, some novel methods are proposed that focus on the naming rules of the generation process and some other special naming features of attributes in databases. These methods are implemented in the COMA CE framework. It is an open source [20] ontology and schema matching tool, and gave good results on OAEI evaluation campaigns [21]. Using the framework of COMA, first the proposed similarity measures were implemented, then complex matchers aggregated these measures; and at the top, “workflows” were defined. COMA supports the use of synonyms and abbreviations, but it is not transparent how the built-in methods use them. Hence, a new implementation was carried out in the framework that allows one to use some features for computing the proposed similarity measures, such as graph parts (parents, children), data types, synonyms and abbreviations. In the following, the proposed methods will be grouped in terms of their application, and later the experiments used to test their performance will also be grouped in the same way.

A. POS Printing Data Integration

In this case, two kinds of methods are proposed for improving the accuracy of ontology alignment.

1. **STRAT1**: In this similarity measure method, the rules of ontology generation are taken into account. The similarity between any two concepts is calculated by separating the parent class names. This can be done easily, because predefined delimiters were used at the naming using class and attribute names in the ontology generation process. After this separation, when comparing concepts, we will only use the attribute name, and a normal string matching method is applied (edit distance based). In addition, this method and the others as well, contain a type comparison related to ontology: when different typed concepts are compared (class to not class, data type property to not data type property) then only a predefined minimal score is given as the similarity value. It should be mentioned here that this similarity measure is suitable only for concept names in the same language. Hence, our algorithm is the following:
    - Compare the type of the concepts (class, properties)
        - if the types don’t match, then a predefined low constant is the similarity value
    - Create the child names using the predefined delimiters
    - Use string matching method (Edit Distance-based) to determine the similarity value
2. **STRAT2:** This case handles issues when the names of the database tables or attributes do not follow the conventions, so the names can be written in a mixed language perhaps without any conventional separation like uppercase or delimiters. For example, “akciotype” or “aruhaztype” may be mentioned, which mean ‘sale type’ and ‘store type’. Here we solve the problem of accents too, for the Hungarian case. This method contains the following steps:
    - First, compare the type of the concepts (class, properties)
    - Second, the words of global ontology are converted into their non-accented form.
    - Third, the words of a concept in the global ontology written in different languages are collected in a set. However, predefined stop words are not considered.
    - Fourth, the multilingual word set of the global ontology is fitted to the local ontology names, with some restrictions. Only those matched characters are summed for which the length of fitting is at least 2. One character length fitting is not considered.

    The next pseudo code snippet tries to illustrate the essence of this kind of similarity:

```
function getSimilarity (concept1, concept2)
if type(concept1)!=type(concept2)
        return lowValue
name1=toNonAccent(getNames (concept1))
name2=toNonAccent(getNamesMultiLanguage(concept2))
tokens2=tokeniseWithoutStopwords(name2)
tokens2=addSynonyms(tokens2)
foreach T2 in tokens2
        sim=FindTokenGetSimilarityWithGapModel(T2, name1)
        totalSim=totalSim=sim;
    return totalSim
```

    In this way, the matching process will consider cases where concept definitions contain mixed language expressions or mixed word order.
3. **STRAT3**: The third strategy is a combination of the previous strategies. The combination function simply takes the average of the scores.

B. Event Data Integration

This integration application allowed more improvement possibilities in ontology matching. The reason for this may be that there were 16 different alignments, and this number is sufficient for gaining more experience. One central idea here based on some investigations concerning the child concepts of a class. Graph-based similarity measures have been used since many years [11][12], and they are based on two basic ideas:

- Top-down: if two classes are similar, then it is more likely that among the child concepts there will also be more or less similar ones.
- Bottom-up: if among the child concepts there are many similar concepts, then it is more likely that parent concepts are similar as well.

We propose to fuse these two ideas to get a new method:

1. Beginning with leaves, the similarities of parent concepts are computed by aggregating similarities of child concepts.
2. In the next step, the similarity values of child concepts are computed by aggregating their string similarity (or another one) with the similarity of the parent concept.

For ontology matching, the following similarities were implemented:

- S1: edit distance-based similarity with affine gaps
- S2: similar to S1, but if the name of the concept contains the name of the parent, then the parent name is cropped, and just the names of children are compared.
- S3: only child names are compared (if the name contains the name of parent, then it is cropped). Names are then tokenized by delimiters or uppercase letters. The two word series of local and global concepts are then compared, and the maximal similarity is kept. During this matching process, synonym substitution is also applied.
- SP1: Computing a similarity for classes that have children by comparing children. This similarity is

computed by averaging the similarity values of children, which are above a predefined threshold.

- SP2: Similarities of parent classes are computed by aggregating their own similarity values with the similarity of child concepts.
- SP3: Similar to SP2, but the similarity of child concepts is computed by averaging the maximal similarities for each child of the first parent concept. The next 'code' snippet illustrates this similarity:

```
children1=parent1.getChildren()
children2=parent2.getChildren()
for each Ch1 in chidren1
    val=getMaxSimilarityToChildren(Ch1, children2)
    if val>lowThreshold
            totalChSim=TotalChSim+val
            N=N+1
totalChSim=totalChSim/N
aggregatedParentCildSim=nameSim(parent1,parent2)+
                              totalChSim*weightOfChildren
```

- SC1: For child concepts, their previously computed similarity values are aggregated with the similarity value computed for the parent.
- SC2: similar to SC1, but this is computed only for same typed pairs (e.g., object property pair).

Using the above-mentioned similarities, the following matching strategies were developed:

1. **STRAT-GR1**: S1 and SP1 are combined.
2. **STRAT-GR2**: S1, SP2, and SC1 are combined.
3. **STRAT-GR3**: minimal score of STRAT-GR1 and STRAT-GR2
4. **STRAT-GR4**: S3, SP3, and SC2 are combined.

V. EXPERIMENTS

Our experiments were performed on the ontology pairs described above. The reference similarity measures applied were complex strategy matchers of COMA:

- **ComaOpt**: This matcher takes into account graph-based similarities (leaves, parents), path similarities, as well as Levenshtein distance-based string similarities of names.
- **COMA**: This matcher combines name similarities, considers the types of concepts and data, graph-based metrics (leaves, parents, siblings) and path.

The baseline alignments were determined by two different people, and then a decision was made about which ones should be kept and which ones were inaccurate. Some control queries were also defined in our data integration framework to test whether the alignments were suitable. To determine alignments in the POS Printing application, a Java implementation using Alignment API was developed, since here complex alignments were used, described by a "level 2" EDOAL alignment RDF file. In the second experimental set-up, the GUI of COMA was applied to create reference alignments, and "level 0" alignment descriptions were created. In this Event Data integration problem set, two kinds of global ontology were considered. One was the complete ontology generated from Java code of the Google Calendar client API, and the other was a reduced one containing just events and their properties. The precision of alignments found were measured in terms of the precision, recall, and F-measure, which are commonly used metrics in alignment evaluation (see for this for example Ontology Alignment Evaluation Initiative, OAEI on Internet). Precision means the fraction of retrieved alignments that are correct, recall is the fraction of correct alignments that are found, and F-measure combines precision and recall by a harmonic mean. However, for the Event Data integration task just the F-measure will be presented for reasons of space.

VI. RESULTS

The results of our evaluations are divided into two experiments, and will be discussed separately below.

A. POS Printing Data Integration

This integration task contained a generated local ontology, and a global one created by hand for this area, as described earlier. The results of the proposed matching strategies are listed in Table I.

TABLE I. RESULTS OF THE PROPOSED METHODS FOR THE POS PRINTING TASK.

Matching strategy	F-Measure	Precision	Recall
ComaOpt	0.0694	0.0735	0.0658
STRAT1	0.2209	0.2069	0.2368
STRAT2	0.3681	0.3448	*0.3947*
STRAT3	*0.4444*	*0.5085*	*0.3947*

As can be seen, the performance of the reference alignment (ComaOpt) is rather low, so it is not surprising that using some additional knowledge about the structure of generated ontologies, and exploiting the fact that the global ontology is labeled in two languages can greatly increase its performance results. It is also seen that the STRAT2 strategy has a better F-measure value than that for STRAT1. The STRAT3 matching strategy, which is a combination of STRAT1 and STRAT2, raises the precision value. This means that in this application, handling the mixed language and mixed word order effects (STRAT2) is more important than handling the effect of ontology generation rules (STRAT1). However, the most robust solution is a combination of them (the precision of STRAT3 is roughly the sum of those of STRAT1 and STRAT2).

B. Event Data Integration

The results of the Event Data Integration problem set are summarized in tables II and III. Table II contains F-measures of alignments corresponding to local ontologies and the full generated global ontology, while Table III contains those for the restricted global ontology.

Here, both the global and local ontologies are generated, so the naming rules are similar. This means in practice that comparing just the words of the generated names will boost the performance, but not so dramatically as in the previous use case.

Examining Table II first, we see that STRAT-GR1 and STRAT-GR3 can greatly improve the performance in some cases, and it is very interesting that STRAT-GR2 can give a fairly positive result when all the other measures give zero. On average, the STRAT-GR1 and STRAT-GR3 are the most promising methods, while STRAT-GR2 and STRAT-GR4 currently like the COMA baseline ones.

Table III also shows performance improvements in most cases, but these improvements are spread over the proposed methods, and there is no method that is an absolute winner. On average, however, the order of the best is different in Table II: the best one becomes STRAT-GR3, while STRAT-GR1 is the next best one. STRAT-GR2 is also better than the baseline cases.

It should be mentioned again that STRAT-GR1 is a complex matcher that combines string similarity with a one-step child-parent rescoring method, while STRAT-GR3 is similar, but it aggregates the scores of parent similarities with the similarities of children in two steps in a different way. We see that taking into consideration the similarity of descendants is important in most cases when concepts are compared. Other improvements could be achieved by including knowledge in the similarity measures concerning the ontology generation process.

VII. Summary and Future Work

Here, novel similarity methods were proposed for a special case of ontology matching; namely, when typically generated ontologies are the targets of the alignment process. To demonstrate the validity of our concept, experiments were also carried out to verify improvements in the performance for each method applied.

In the future, we plan to create a graphical user interface that supports the notation of complex alignments as well. Moreover, an automatic matcher is planned to help find these complex relations (e.g., inverse, composition, restriction). To evaluate the precision of automatic ontology matching in this case, a modified evaluation process, which takes into account the type of complex alignments will also be needed. Another interesting area might be to adaptively improve complex matching models by tuning their parameters, which means applying and customizing adaptive learning techniques to our particular case.

VIII. Acknowledgements

This work has been partially supported by the European Union and the European Social Fund through project FuturICT.hu (grant no.: TAMOP-4.2.2.C-11/1/KONV-2012-0013). Also, the work has been supported by the European Union, co-financed by the European Regional Development Fund (grant no.: GOP-1.1.1-11-2011-0052).

IX. References

[1] A. K. Alasoud, “A Multi-Matching Technique for Combining Similarity Measures in Ontology Integration,” Phd Thesis, Concordia University Montréal, Québec, Canada, 2009.

[2] P. Shvaiko and J. Euzenat, “Ontology Matching: State of the Art and Future Challenges,” IEEE Trans. Knowl. Data Eng., vol. 25, no. 1, IEEE Educational Activities Department Piscataway, NJ, USA, 2013, pp. 158–176.

[3] D. H. Ngo and Z. Bellahsene, “YAM++ - Results for OAEI 2012,” in International Semantic Web Conference, United States, 2012.

[4] D. Engmann and S. Maßmann, “Instance Matching with COMA++,” in BTW 2007 Workshop: Model Management und Metadaten-Verwaltung, Aachen, Germany, 2007, pp. 28–37.

[5] Y. Kalfoglou and B. Hu, “CMS: CROSI Mapping System - Results of the 2005 Ontology Alignment Contest,” K-Cap’05 Integrating Ontologies workshop, Banff, Canada,, 2005, pp. 77–85.

[6] Q. Ji, P. Haase, and G. Qi, “Combination of Similarity Measures in Ontology Matching using the OWA Operator,” in: Proceedings of the 12th International Conference on Information Processing and Management of Uncertainty in Knowledge-Base Systems, june 22-27, Malaga, 2008.

[7] N. Choi, I.-Y. Song, and H. Han, “A survey on ontology mapping,” Journal: SIGMOD Record, vol. 35, no. 3, ACM New York, NY, USA, 2006, pp. 34–41.

[8] G. Stoilos, G. Stamou, and S. Kollias, “A string metric for ontology alignment,” in Proceedings of the 4th international conference on The Semantic Web, Berlin, Heidelberg, 2005, pp. 624–637.

[9] M. Espinoza, A. Gómez-Pérez, and E. Mena, “LabelTranslator - a tool to automatically localize an ontology,” in Proceedings of the 5th European semantic web conference on The semantic web: research and applications, Berlin, Heidelberg, 2008, pp. 792–796.

[10] J. Euzenat and P. Shvaiko, Ontology Matching, (book) 1st ed. Springer Publishing Company, Incorporated, 2010.

[11] N. F. Noy and M. A. Musen, “Anchor-PROMPT: Using Non-Local Context for Semantic Matching,” in Proceedings of the workshop on ontologies and information sharing at the international joint conference on artificial inteligence (IJCAI), Seattle, Washington, USA , 2001, pp. 63–70.

[12] M. H. Seddiqui and M. Aono, “An efficient and scalable algorithm for segmented alignment of ontologies of arbitrary size,” Web Semantics: Science, Services and Agents on the World Wide Web, vol. 7, no. 4, 2009, pp. 344 – 356.

[13] S. A. McIlraith and D. L. Martin, “Bringing semantics to Web services,” Intelligent Systems, IEEE, vol. 18, no. 1, pp., IEEE Educational Activities Department Piscataway, NJ, USA, 90 –93, Feb. 2003.

[14] O.. Curé, M. Lamolle, and C. L. Duc, “Ontology Based Data Integration Over Document and Column Family Oriented NoSQL stores,” in The 7th International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS2011), Bonn, Germany, 2011.

[15] C. Kavitha and G. S. Sadasivam, “Ontology Based Semantic Integration of Heterogeneous Databases,” in European Journal of Scientific Research, 2011, vol. Vol.64 No.1, pp. 115–122.

[16] http://oaei.ontologymatching.org/ (2013)

[17] C. Bizer, “D2RQ - treating non-RDF databases as virtual RDF graphs,” in In Proceedings of the 3rd International Semantic Web Conference (ISWC2004, poster presentation), Hiroshima, Japan, 2004.

[18] http://qdox.codehaus.org/ (2014)

[19] A. Banhalmi, D. Paczolay, A. Z. Vegh, G. Antal, and V. Bilicki, “Development of a Novel Semantic-Based System Integration Framework,” in Engineering of Computer Based Systems (ECBS-EERC), 2013 3rd Eastern European Regional Conference on the Engineering of Computer Based Systems, Budapest, Hungary, 2013, pp. 18–24.

[20] D. Aumueller, H.-H. Do, S. Massmann, and E. Rahm, “Schema and ontology matching with COMA++,” in Proceedings of the 2005 ACM SIGMOD international conference on Management of data, New York, NY, USA, 2005, pp. 906–908.

[21] J. Euzenat at al., “Results of the ontology alignment evaluation initiative 2011,” in Proc. 6th ISWC workshop on ontology matching (OM), Bonn (DE), pages 85–110, 2011.

TABLE II. RESULTS OF REFERENCE AND PROPOSED MATCHING METHODS IN TERMS OF THE F-MEASURE, WHEN A LARGE GLOBAL ONTOLOGY IS USED IN THE EVENT INTEGRATION USE CASE.

Application/Matching strategy	**Coma-Opt**	**COMA**	**STRAT-GR 1**	**STRAT-GR 2**	**STRAT-GR 3**	**STRAT-GR 4**
Basic-php-events-lister2.04	0.25	0.25	*0.29*	0.20	0.26	*0.29*
calendar	*0.21*	0.25	0.00	0.13	0.00	0.00
calendar_v2.0_en	0.14	0.17	*0.27*	0.20	0.21	0.23
calendar_ws	0.00	0.00	0.00	*0.25*	0.00	0.00
calendarix_0_8_20080808	0.13	0.05	0.13	0.17	0.20	0.11
calendartechnique-2.0.2RC4	0.19	0.23	*0.30*	0.12	*0.32*	0.07
cmappCalendar_1.1	0.19	0.19	*0.82*	0.13	0.75	0.59
Fullcalendar	0.00	0.00	0.00	*0.29*	0.00	0.00
luxcal273	*0.24*	0.17	0.20	0.10	0.17	0.11
maian_events	0.21	0.24	*0.31*	0.19	0.25	0.20
mapcal-0.2.1	*0.36*	*0.36*	0.35	0.27	0.30	*0.36*
openbookings.org_v0.6.4b	0.00	0.00	0.00	*0.35*	0.00	0.00
PHPCalendar.Basic.2.3	0.19	0.07	0.00	0.00	0.00	0.00
supercali-1.0.7	0.15	0.20	0.15	0.17	*0.21*	0.10
vcalendar_php_1.1.5	0.22	0.25	*0.53*	0.20	0.43	0.17
webical-0.4.1	0.43	0.46	0.45	0.39	*0.60*	*0.62*
Average	0.18	0.18	*0.24*	0.20	*0.23*	0.18

TABLE III. RESULTS OF REFERENCE AND PROPOSED MATCHING METHODS IN TERMS OF THE F-MEASURE, WHEN A RESTRICTED GLOBAL ONTOLOGY IS USED IN THE EVENT INTEGRATION USE CASE.

Application/Matching strategy	**Coma-Opt**	**COMA**	**STRAT-GR 1**	**STRAT-GR 2**	**STRAT-GR 3**	**STRAT-GR 4**
Basic-php-events-lister2.04	0.40	0.36	0.44	0.46	0.44	*0.54*
calendar	*0.33*	0.00	0.17	0.27	*0.33*	0.00
calendar_v2.0_en	0.40	0.40	*0.53*	0.44	0.44	*0.53*
calendar_ws	0.20	0.00	0.20	*0.67*	0.20	0.20
calendarix_0_8_20080808	0.30	0.27	0.39	0.33	0.39	0.22
calendartechnique-2.0.2RC4	0.27	0.28	*0.40*	0.30	*0.40*	0.15
cmappCalendar_1.1	0.56	0.56	*0.88*	0.47	*0.88*	0.80
Fullcalendar	0.60	0.00	0.55	*0.71*	0.55	0.40
luxcal273	*0.52*	0.40	0.46	0.35	0.45	0.44
maian_events	0.53	0.40	0.63	0.57	0.63	*0.67*
mapcal-0.2.1	0.61	*0.64*	0.50	0.42	0.48	0.57
openbookings.org_v0.6.4b	0.00	0.00	0.00	*0.24*	0.00	0.00
PHPCalendar.Basic.2.3	0.33	0.32	*0.53*	0.33	0.44	0.25
supercali-1.0.7	0.67	0.67	0.53	0.59	*0.71*	0.53
vcalendar_php_1.1.5	0.64	0.55	0.72	0.48	*0.73*	0.50
webical-0.4.1	0.56	0.58	0.65	0.67	*0.73*	0.69
Average	0.43	0.34	0.47	0.46	*0.49*	0.40

Improving Large Image Viewing in the Crisis Management System SécuRéVi

Mehdi Tahan, Jean Vareille,
Laurent Nana
Laboratoire des Sciences et Techniques de l'Information, de la Communication et de la Connaissance
Brest, France
e-mail: {Mehdi.Tahan, Jean.Vareille, Laurent.Nana}@univ-brest.fr

Olivier Danjean
Inovadys
Brest, France
e-mail: Olivier.Danjean@inovadys.com

Hervé Mahoudo, Gilles Cloarec
SDIS29 & EDF
Brest, France
e-mail: Herve.Mahoudo@sdis29.fr, Cloarec_Gilles@hotmail.com

***Abstract*— This paper deals with the processing and integration of images in SécuRéVi platform, a crisis management system, with an emphasis on large images whose handling leads to specific difficulties. Indeed, in the domain of crisis management, images are key elements for understanding and taking proper decisions. Images of different nature (map, aerial view, 360 ° view, etc.) are helpful and adequate solutions that need to be provided in case of sinister. After describing the SécuRéVi platform, we shortly present its images processing approach and establish some design constraints and criteria for the development of a large image viewer. Then, we present our proposal for improving large image handling and its implementation in the SécuRéVi platform. The paper ends by conclusion and future works.**

Keywords-crisis management; image processing; large image; image segmentation

I. INTRODUCTION

Crisis management requires a good understanding of the sinister environment for adopting necessary attitudes for the resolution of incidents. Different prevention measures are provided for this purpose, such as emergency planning.

The SécuRéVi [1] platform is inscribed in this process. It is based on the Global Safety Plan (GSP), which is a knowledge base aimed at gathering all the information needed by preventive approaches dedicated to crisis management. The understanding of a situation is a major challenge for crisis resolution and images are key elements for such understanding (like in geospatial imagery, scientific visualization or immersive application [2]). It is therefore important for a crisis management tool, to provide adequate solutions for images handling. In the context of crisis management, images can come from various sources (such as satellite [3], camera [4] or industrial plan) and are incorporated before or during crisis management. They are extracted from camera shots or shots assembling [5]. The size of the used images goes from tens to hundreds of millions of pixels. The image generation systems offer increasingly higher resolutions, but often, visualization tools do not allow their full restitution because the screen resolution is most often lower than that of camera sensors. Given the large size of images generated, the use of conventional tools quickly becomes problematic and they can't be used by any type of computer [6]. Therefore, we decided to develop a tool allowing a better use of these images.

We start by describing the SécuRéVi Platform (section 2). Then, we show the importance of images for crisis management and explain their handling in the SécuRéVi platform (section 3). Thereafter, we present our proposal for image processing and precise the key elements for a successful implementation (section 4). Afterwards, we describe the implementation of this process and its evolution to obtain better performance (section 5). Finally, we conclude and propose some lines of thought for future works (section 6).

II. SÉCURÉVI PLATFORM

The SécuRéVi Platform [7] is a platform dedicated to communication and understanding, as well as decision making and the workmanship (see Figure. 1). It allows different actors from various professions to communicate and collaborate for a crisis resolution. It is based on the GSP integrated into a Geographic Information System (GIS), as well as various sources of external data (measurement tools, camera, business oriented computer tools, etc.).

The whole interacts with a monitoring and events forecasting system (accident, incident, etc.) with commitment of resources (staff and equipment) in which each object, event or staff, can index data (e.g., description sheet, procedure, user manual, etc.) .

SécuRéVi allows an approach in both space and time. It provides real time monitoring which makes it possible to create recording for replays that will feed the "lessons learned" database associated (feedback). Finally, in the platform, a 2D world can be associated with a 3D world, and this makes it possible to go from the virtual world to the real world and inversely.

SécuRéVi works with a standard computer which can be enriched with other tools (video projector, interactive whiteboard, etc.).

ISBN: 978-1-61208-342-1

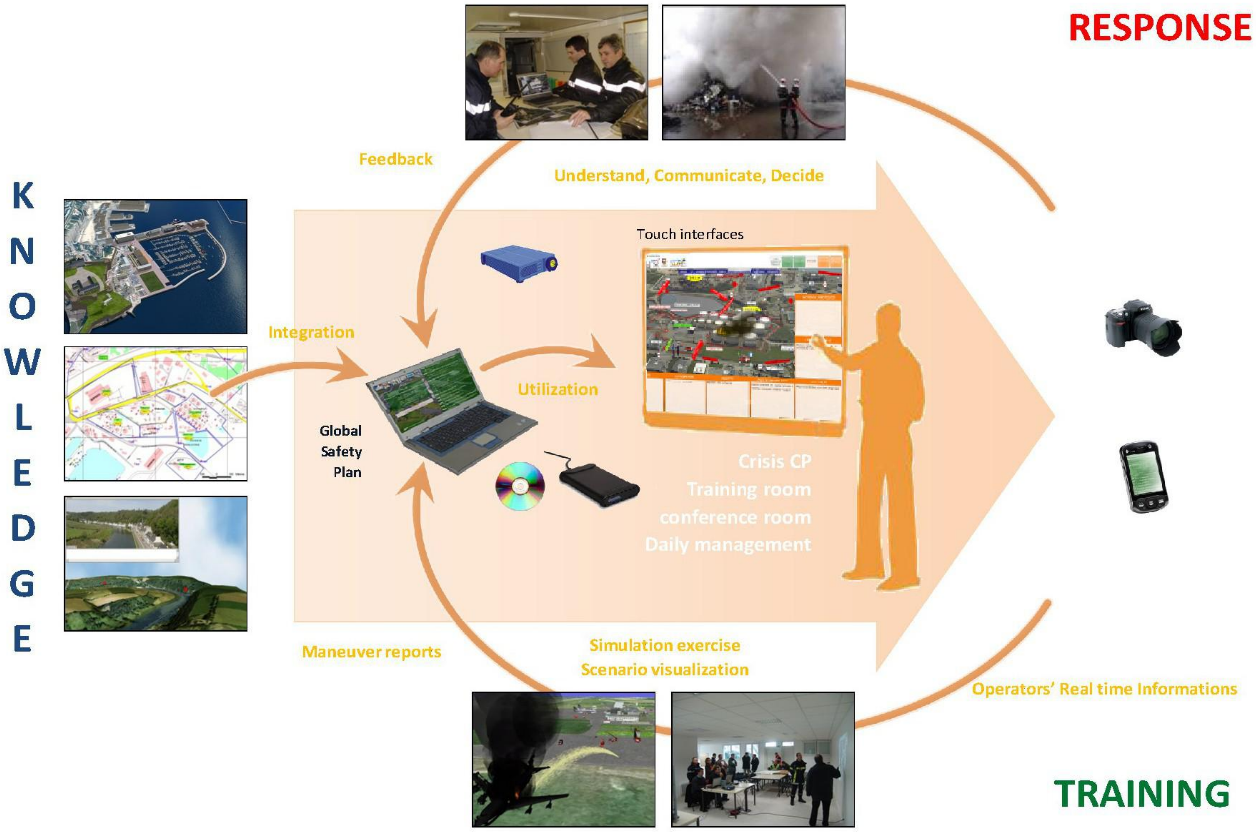


Figure 1. The SécuRéVi platform (crisis management system).

It is used during crisis management operations, during which the in-situ data and GSP allow a better understanding of the context. Any intervention concludes with a feedback that can enrich the knowledge base for future operations. Training will take advantage of the platform which makes it possible to use virtual information while maintaining realistic working framework.

The initial concepts and methodologies of the SécuRéVi platform were developed at the Research and Development Expertise Service (Direction Expertise Recherche et Développement, DERD) of the Departmental Office of Fire and Rescue of Finistère (Service Départemental d'Incendie et de Secours du Finistère, SDIS29) by Colonel Hervé Mahoudo and engineer Gilles Cloarec of SDIS29, and engineer Olivier Danjean, head of the company INOVADYS.

This platform is used by Colonel Hervé Mahoudo and Gilles Cloarec for courses offered at the National School of Firefighters Officers (ENSOSP, Ministry of Interior), in the following areas:

- Classified Installations for the Protection of the Environment
- Chemical hazards (internships RCH4)
- Emergency Planning [8],
- Crisis management.

The platform and the concepts it includes are also the subject of education provided by Colonel Hervé Mahoudo and Gilles Cloarec, during courses at the University of Western Brittany (UBO), at the University of South Brittany (UBS), at the University of Bordeaux (in the QHSE field) and at the University of Rennes 2 (in master of GIS).

On-site implementation consists in achieving the institution's GSP, installing tools for management, monitoring and forecasting of situations and deploying hardware interfaces.

SécuRéVi constitutes a multidisciplinary and multiservice digital knowledge database for daily operations, training, intervention and communication. It is usable by the rescue operations commandant, the leader of internal operations, site safety manager, trainers and all other internal persons authorized by the company.

In the next part, we show the importance of images for crisis management and explain their handling in the SécuRéVi platform.

III. Importance of Images in Crisis Management and Their Handling in SécuRéVi Platform

The decision making in crisis management needs different types and amount of information depending on the complexity of the situation. Usually, a rapid-mapping is the first task needed to understand the overall situation. More the

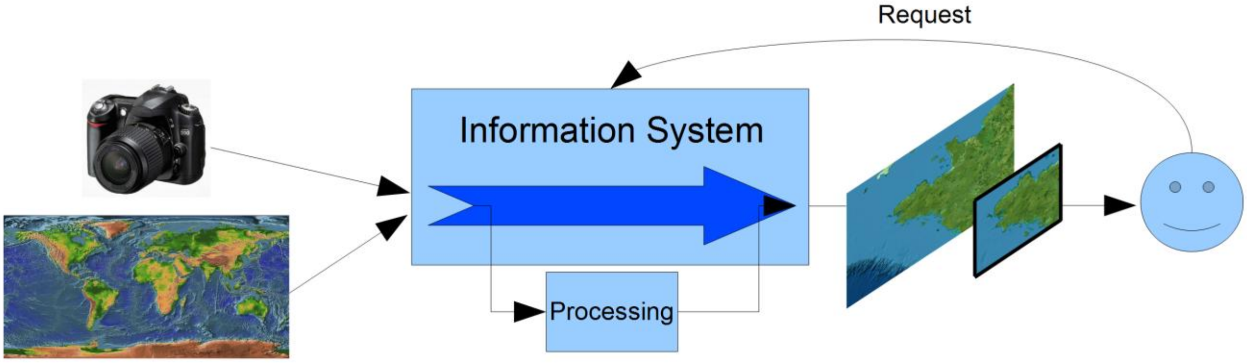


Figure 2. From integration to viewing.

decision process progresses, more the need of detailed information is important.

Images are certainly the most appropriate media for these tasks. For example, satellite images can provide a rapid and overall comprehension of a site, plans allow to see the structure of the building and 360° view help to understand the reality of a room including its content. The use of high definition images improves these tasks.

Therefore, the SécuRéVi platform allows intensive use of visual representations as images: 8 large images (as satellite views, plans and aerial views) and 130 pictures of 360° views were required for a supermarket whose surface is about 21 000 m².

The general working principle of image handling in SécuRéVi is shown in Figure 2. Image sources range from camera to GIS and the formats of the images generated are varied. JPEG has established itself in the field of digital photography while there are several commonly used file formats, such as ECW format, in the domain of GIS. The integration of original image files in the Information System (IS) was the process in place before the work presented in this paper. In this initial process, the user connects to the IS and displays the file directly. The IS is typically local to the machine but may be incorporated within a remote storage area. The "processing" part of Figure 2 represents the processing incorporated for large images, which will be discussed in the next section.

IV. Proposal of a tool for large images handling

The hope in decision making in crisis management is to get all the data instantaneously. Obviously, this is not possible in practice due to hardware and software constraints. Moreover, the handling of large images adds new constraints (storage, processing).

The design challenge is then to manage the contradiction between the hope and the reality and to find an acceptable solution in terms of delay, despite the huge size of images.

In this section, we start by presenting the design constraints of the tool, then we precise our design criteria. The solution proposed is then described, as well as its implementation.

A. Design constraints

In the context of crisis management, the machines used for image visualization have various processing capacities. A machine with a limited performance x86 processor (as Intel Atom), 1 GB of RAM and a hard drive's capacity of 200 GB was defined as the minimum hardware requirement. A 400 Megapixel JPEG (common for a GIS document) uses several GB of RAM and becomes difficult to use on this type of machine. However, on a machine with sufficient memory, the opening time of a file remains problematic.

The tool must accept files from different sources and optimize the use of available resources.

There are different means, like the conversion to a specific format. For example, the use of a wavelet image compression format has been considered as one of the possibilities. It has not been adopted because the conversion with existing tools needs too many resources.

B. Design criteria

For our process, we define an Ideal Final Result (IFR) [9] as a target. This IFR needs to take into account user and computer. We define 3 measuring elements: display time, memory usage and CPU time. These elements need to be as lower as possible. The display time is the time between the user request and the display of the image on the screen. A value less than 100 ms can be considered as a coherent value with the human physiology. We determine the memory usage value like the product between the total number of pixels of the screen and the color depth. The CPU time needs to be near 0%.

Other criteria have been taken into account, including:

- no disturbance of the working of the other parts of SécuRéVi,
- the friendliness when using the tool proposed.

C. Proposed solution

The proposed image processing tool has two parts, one for the pre-processing of images and the other for their visualization. The pre-processing is inserted at the beginning of the restitution chain. Visualization transparently replaced the tool previously used. So, the integration of the new tool does not disturb the working of the other modules of the SécuRéVi platform.

The pre-processing aims at optimizing the use of resources. It relies on the use of cutting tiles which is a known technique with different implementations [10]. It is popularized by online mapping services [11]. It minimizes the memory use: only visible tiles are loaded (which are used to generate a new image by assembling and extracting the final visualization). However, when viewing the whole image, the entire tile need to be loaded in memory and the video card need to downscale it before the display. So in this case, the gain is zero in comparison to the initial viewer. In order to improve the gain, we have integrated the management of multi-resolution into the process. This allows to display only the paving tiles whose resolution is directly above the output's display resolution (see Figure. 3). Data formats such as JPEG2000 [12] allow this type of cutting but neither their production nor their return meet our requirements as far as memory use is concerned.

From the pixel array, we generate new images (tiles) organized in predefined zoom level. This data tree is packaged in an archive. The viewer selects the tiles to be displayed according to their level of zoom and resolution. It loads tiles on the fly and releases the useless ones when necessary.

The resolution of a tile was defined empirically: the longest side must have a size of 512 pixels, the value of the second one keeps the ratio of the original picture. Different values were tested. Multiples of 2 show the best performance in terms of processing, but less than 512 pixels value generates too many files (which can be problematic in a FAT32 file system) while a higher value will not make optimal use of memory visualization module (in the "worst case" a display resolution lower than 512x512 pixels will require an image of 2048 x 2048 pixels where our choice divides by 4 the resolution required). In practice, an image of 12100 per 6050 pixels (and weighted 10,8MB in Jpeg), needs the generation of 510 tiles (for a weight of 23,5 MB).

Regarding performance, the software (processing and visualization) needs less than one hundred MB (before it was necessary to have several GB). The CPU load is maximum during processing but minimal during playback. When the processing is done beforehand, the minimum required configuration can be used in good conditions. Finally, one of the consequences of pre-processing is the almost instantaneous loading compared with the tens of seconds of the old loading method (on the minimum performance machine).

V. IMPLEMENTATION

Two versions of the image processing software have been developed. The first version has been tested and integrated into the SécuRéVi Platform. The second version is an optimized version of the first one. We start by presenting the first version software and tackle its testing, then we present its second version.

A. *First version of the image processing software*

1) *Principles and working*

In this version, before any use of an image, its entire pre-processing is done and an archive of the result is created as a single file. This use matches a company specific request: generate spherical view. Generating a spherical view consists in assembling photographic images each having a different rotation around a central point. These views are rectified and merged according to specific criteria. The integration of the initial photographic views directly into the IS is maladjusted and the addition of the pre-processing is not detrimental. In contrast, pre-processing does not match the dynamic integration of new data during use (for example, for a crisis management). This problem was solved by storing the images/archives on external transportable disks, such as to be able to use them on the intervention site. Nowadays, data transport by external device is less used and replaced by an access through the network. As a consequence, it is no more interesting to use a single archive. Indeed, the entire file must be downloaded before viewing. Since the archive is heavier than the original file, the interest is therefore diminished. For these reasons, two main changes have been done:

- Archives have been replaced by simple data trees. This change limits the network use negative impact (the amount of data transferred is smaller).
- The pre-processing has been replaced by a processing on request: the pre-processing is only called if the tiles required are not available.

The generation on request allows to optimize storage: only image sections needed are generated and stored. This feature opens a new possibility for analyze like the detection of point of interest [13].

For example, more than 3,5 GB were generated for all of the 360° views (with the pre-processing process) and about 80MB for the large images (with the on request process) for the supermarket site mentioned in section 3.

2) *Testing and results*

The table below (Table 1) is used to compare the processings. The measurements were performed on a machine equipped with a i5-2520 M processor and 8 GB of RAM and the processed image resolution was 16384 x 8192.

After several on-site deployments, we have chosen to accept only JPEG format as input. Third-party software can often export to this file format, otherwise screen printing can be used (at the expense of resolution).This allows to exploit their expertise in their respective fields but also allows great flexibility of workflow for users, and this is very useful in dynamically changing settings [14] as it is the case in a crisis context.

However, we encounter limitations due to software that allow the export of pictures. High resolutions (over 20 000 x 20 000 pixels) are poorly supported. JPEG format accepts a

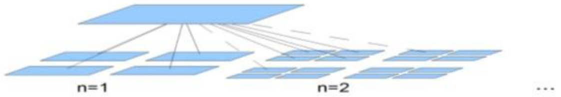


Figure 3. Multi-resolution tiles organization.

maximum resolution of 65535 in width and height. Format as JPEG2000 would achieve higher resolutions as input. The tiles format seems not to need to evolve: the JPEG covers all needs and viewing, despite the loss of information during processing (rasterization, compression, etc.). The data produced were not intended to return to their original business domain so, the loss of information doesn't have any consequence.

One of the services made possible by the use of the network is the unsupervised distributed computing. One of the conditions of its implementation would be a reasonable network throughput. The current algorithm is not optimized for this type of use. However, this operation is already operational but not optimal.

B. Optimized version

The optimized version generates a tile by the assembly of 4 tiles of the lower level (the first method uses the original image for every tile). It reduces the time needed by 20 compared with the original pre-processing method. The limit is the need to generate all of the tiles of the lower levels before visualizing the highest level tile. It is not optimized for our on request process but we are working on the integration of intermediate operation in order to reach better performance.

VI. CONCLUSION AND FUTURE WORK

In this paper, we have proposed a solution for the handling of large images in a crisis context. The solution proposed has been integrated in the crisis management platform SécuRéVi. Existing tools such as JPEG2000, were not adapted to our needs due to some constraints such as restricted resources (memory, CPU rate, display time), dynamic integration of images (on the fly).

The solution proposed has been tested and evaluated according to 3 criteria: image display time, memory usage and CPU time. The result shows that the new software tool has good performances. After the first viewing, the image display time has been divided by twelve and the CPU time is lower in comparison to that of initial tool. The percentage of memory used has decreased from 100% to 14%.

In operation, the major contribution to the business experts is the ability to update the IS by adding image during planning or operation. The new process allows to use new data very quickly and updating data is a very important feature in a crisis management system [15].

Our industrial partners (EDF, Total, Jeddah airport, industrial and port area of Le Havre, etc.) frequently use our viewer and their returns are positive.

An optimization of the tool has been proposed in order to reduce the time needed for tiles generation by a coefficient of 20. We plan to integrate it in the SécuRéVi platform. This will be helpful especially for tasks needing complete generation like 360° view.

It is very important to take into account the evolution of the technological environment (such as a GPU integration [16], the use of an UHD display or new services like indexing files) to respond quickly and efficiently to users requirements. One of the future works is the investigation of the use of innovative methods such as TRIZ to adapt our solution accordingly.

We also plan to study the use of grid computing to fasten the tiles generation process and to formalize the determination of the size of tiles (currently defined empirically).

REFERENCES

[1] M. Tahan, J. Vareille, L. Nana, O. Danjean, H. Mahoudo, and G. Cloarec , "Systèmes Complexes et Plan Global de Secours (P.G.S.): vers un plan global Qualité, Hygiène, Sécurité et Environnement (Q.H.S.E.)" [Complex System and Global Safety Plan (G.S.P) : towards a global plan for Quality, Health, Security and Environment (Q.H.S.E.)], CNRIUT, Tours, France, Jun. 2012.

[2] T. Ni et al., "A Survey of Large High-Resolution Display Technologies, Techniques, and Applications", Alexandria, Virginia, USA, Mar. 2006, pp.223-236

[3] L. Montoya, "Geo-data acquisition through mobile GIS and digital video: an urban disaster management perspective", Environmental Modelling & Software, vol. 18, Mar. 2003, pp.869-876.

[4] N. Tholey et al. , "Utilisation de l'imagerie satellitaire pour la gestion de crise type "risques naturels", inondations et risques littoraux" [Using satellite imagery for crisis management of "natural hazards", flood and coastal risks type], 7iemes journées scientifiques et techniques du CETMEF, Paris, Dec. 2008.

[5] J. Kopf, M. Uyttendaele, O. Deussen, and M. F. Cohen, "Capturing and Viewing Gigapixel Images", ACM Traansactions on Graphics, vol. 26, Jul. 2007, pp.93-102.

[6] E. C. Shenchang, "QuickTime VR – An Image-Based Approach to Virtual Environment Navigation", SIGGRAPH '95 Proceedings of the 22nd annual conference on Computer graphics and interactive techniques, 1995, pp.29-38.

[7] H. Mahoudo, "Ne pas être déconnecté du monde réel" [Not to be disconnected from the real world], Face au risque, vol. 488, Dec. 2012, p.13.

[8] H. Mahoudo and G. Cloarec, "Préparation à l'opération. Conceptions des plans d'établissements répertoriés" [Preparing operation. Design of classified institutes plans], ENSOSP prévision, Tome 6bis, 2011.

[9] G. S. Altshuller, "The innovation algorithm TRIZ: systematic innovation and technical creativity", Technical Innovation Center Inc., 1999.

[10] N. R. Pal and S. K. Pal, "A review on image segmentation techniques", Pattern Recognition, vol.26, no.9, 1993, pp.1277-1294.

[11] C. De Souza Baptista et al. , "On Performance Evaluation of Web GIS Applications", DEXA'05, Database and Expert Systems Applications, Copenhagen, Aug. 2005, pp.497-501.

[12] ISO/IEC, ISO/IEC 15 444-1, "Information Technology - JPEG 2000 Image Coding System", 2000.

[13] J. Laflaquière, Y. Prié, and A. Mille, "Ingénierie des traces numériques d'interaction comme inscriptions de connaissances" [Digital traces engineering of interaction as knowledge inscriptions], 19èmes Journées Francophones d'Ingénierie des Connaissances, Nancy, Jun. 2008, pp.183-195.

[14] W. M. P. Van Der Aalst, and M. Weskez, "Advanced Topics in Workflow Management: Issues, Requirements, and Solutions", Journal of Integrated Design and Process Science, vol. 7, 2003, pp.49-77.

[15] M. Turoff, M. Chumer, B. Vand De Walle, and X. Yao, "The design of a dynamic emergency response management

information system (DERMIS)", Journal of information technology theory and application, vol. 5, 2004, pp.1-35.

[16] B. Fulkerson and S. Soatto, "Really quick shift: Image segmentation on a GPU", European Conference on Computer Vision, vol. 6554, Heraklion, Crete, Greece, Sep. 2010, pp.350-358.

TABLE I. COMPARATIVE PERFORMANCE VISUALIZATION LEVEL

	Initial method		Pre-processsing		Processing on request	
	1st viewing	Next viewing	1st viewing	Next viewing	1st viewing	Next viewing
Display time	6 s	6 s	> 20 min.	< 0.5 s	30 s < t < 70 s	< 0.5 s
Memory usage	580 MB		< 20 MB	80 MB	80 MB	
CPU time	100 %	100 %	100 %	<15%	100 %	< 15 %

Towards Sensor-Aided Multi-View Reconstruction for High Accuracy Applications

Mikhail M. Shashkov, Mauricio Hess-Flores, Shawn Recker, and Kenneth I. Joy
Institute for Data Analysis and Visualization
University of California – Davis
Davis, USA
mmshashkov@ucdavis.edu, mhessf@ucdavis.edu, strecker@ucdavis.edu, kenneth.i.joy@gmail.com

***Abstract*—We present the general idea of a computer vision structure-from-motion framework that makes use of sensor fusion to provide very accurate and efficient multi-view reconstruction results that can capture internal geometry. Given the increased ubiquity and cost-effectiveness of embedding sensors, such as positional sensors, into objects, it has become feasible to fuse such sensor data and camera-acquired data to vastly improve reconstruction quality and enable a number of novel applications for structure-from-motion. Application areas, which require very high accuracy, include medicine, robotics, security, and additive manufacturing (3D printing). Specific examples and initial results are discussed, followed by a discussion on proposed future work.**

Keywords-sensor fusion; embedded sensors; multi-view reconstruction; structure-from-motion; Kinect.

I. INTRODUCTION

In the past few years, there has been a great increase in the amount of sensors that are embedded into every day devices on account of the positive trends in lower costs and miniaturization. For example, consider a modern Android**®** or iOS**®** phone whose internal sensors (Global Positioning System (GPS), camera, gyroscope, magnetometer, accelerometer, proximity, audio, and more) drastically outnumber the bigger and less capable cellular phones of prior generations. This trend extends outside of industry and into research where other common sensors, including radar, sonar, LIDAR, infrared, seismic, and magnetic have become utilized more often.

The ubiquity of such sensors and their data creates a *sensor fusion* problem. Sensor fusion involves combining data acquired from different sources in order to provide more accurate or complete information about the sensed target than if these sources were utilized individually. Fusion is non-trivial, and is a very relevant topic today in fields such as computer vision.

In computer vision, one specific instance of sensor fusion is the Red-Green-Blue-Depth (RGB-D) camera, such as the Microsoft Kinect**®**, which jointly acquires color (RGB) data and depth (D) values for each pixel. The addition of depth freed the Kinect from a certain amount of dependence on analyzing only color to do feature detection, object identification, edge detection, and other fundamental parts of object reconstruction. This boon for research in reconstruction and many other fields culminated in KinectFusion [1], which we describe in the next section. But even the KinectFusion has practical limitations for high-accuracy applications because depth estimates tend to be noisy, and without very accurate filtering, are generally not accurate enough to provide reliable data for up-and-coming applications in medicine, 3D printing and robotics. More traditional methods, like structure-from-motion, space carving and others [2] can be more accurate but are typically less dense. These issues are only exacerbated for additively manufactured objects, which are typically texture-less and mono-colored when produced by current consumer hardware. Figure 1 shows some examples of these objects, including a fully functional ball bearing whose reconstruction would have to be very precise and take into account internal geometry (something the systems discussed cannot do due to occlusions) to maintain functional geometry once reconstructed.

Inspired by the gains achieved from adding depth measurements, we investigate the benefits of using positional sensors to assist in multi-view scene reconstruction. To that end, we present initial results on the development of a generalized framework for 3D scene reconstruction aided by any mix of positional data, such as RGB-D or sonar and photographic imagery. Furthermore, we explore the idea of placing these sensors internally in order to reconstruct internal structure. We show that fusing positional data with traditional images improves the accuracy of camera pose estimation and scene reconstruction, especially when dealing with texture-less or mono-colored objects. This fusion also has the potential to capture internal structure as opposed to standard structure- from-motion approaches. Background and related work is discussed in Section II. Some concrete applications and results will be discussed in Section III. Conclusions and future work will be discussed in Section IV.

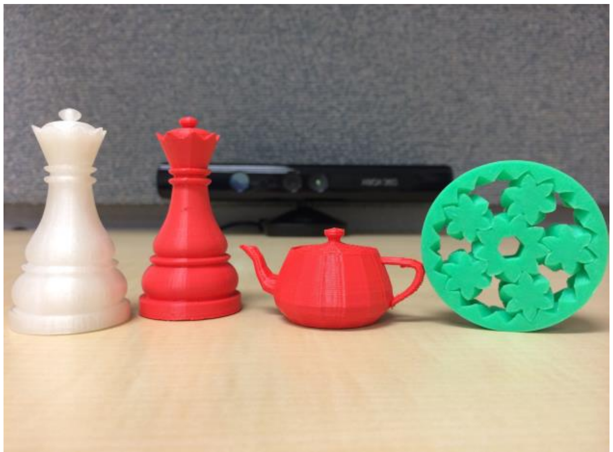


Figure 1. Additively manufactured objects that present challenges.

II. Background and Related Work

To our knowledge, we are the first to propose using internally embedded sensors for multi-view reconstruction. We provide a general background on computer vision and contributions towards scene reconstruction in Section II-A, and discuss recent work on fusing sensing technology with imagery, specifically RGB-D cameras, in Section II-B. We will also discuss recent work on the imaging of internal geometry in Section II-C.

A. General Background of Scene Reconstruction

The broad field of computer vision includes important sub-fields such as object detection, tracking, and the multi-view reconstruction of scenes. The goal of multi-view scene reconstruction is to extract a 3D point cloud representing a scene when given multiple views (such as photographs) of the scene. Detailed analysis and comparisons between methods are available in the literature [2]. Most of these methods seek to create correspondences between views, usually by detecting features and tracking them from view to view. One of the main algorithms used to do this is Scale-Invariant Feature Transform (SIFT) [3]. For an excellent overview of many classical vision algorithms, the reader is referred to Hartley and Zisserman [4].

One drawback of current computer vision methods is that many are based on the mathematical optimization of initial parameter estimates to achieve accurate results. Though such optimization is provably necessary, such as in the case of the well-known *bundle adjustment* [5] in structure-from-motion, the final accuracy is simply not enough for applications that require an extreme amount of accuracy. Furthermore, the density of these reconstructions often leaves something to be desired.

B. RGB-D Cameras

To alleviate the density problem, there has been interest in utilizing depth sensing technologies for object reconstruction for a long time [6], but it is only recently that the technology has become very affordable and easy to use with the release of the Microsoft Kinect® in late 2010. With it, came a plethora of reconstructions of people [7] and indoor environments [8]. One of the biggest successes is KinectFusion, which fuses depth data and RGB data from a movable Kinect in real time to create a dense scene reconstruction as the user moves through the scene. Given its ubiquity and success, we will further detail the KinectFusion algorithm [1][9], since this is the main algorithm we want to challenge as far as reconstruction density and accuracy for our intended applications. The main goal of KinectFusion is to fuse depth data acquired from a Kinect sensor into a single, global surface model of the viewed scene, in real-time. Additionally, 6DOF sensor pose is simultaneously obtained by tracking the live depth frame relative to the global model using a coarse-to-fine Iterative Closest Point (ICP) algorithm [20].

The KinectFusion algorithm can be considered an upgrade to previous ‘monocular Simultaneous Localization and Mapping (SLAM)’ systems [21], the most successful being the Parallel Tracking and Mapping (PTAM) system [10]. The main drawback of those systems is that they are optimized for efficient camera tracking, but produced only sparse point cloud models. Even in novel systems, which combine PTAM’s camera tracking capability with dense surface reconstruction methods (such as described in [1]) in order to enable better occlusion prediction and surface interaction [11][12], dense scene reconstruction in real-time remains a challenge. Results are still highly dependent on factors such as camera motion and scene illumination.

KinectFusion has been proven to work well for situations with a dynamic element involved: either the objects in the scene or the camera itself is moving. In our applications, we're more interested in acquiring a very high level of detail, even from a completely static setup. For instance, the KinectFusion algorithm relies on bilateral filtering on the initial depth maps, for noise removal. Though very helpful towards the original algorithm, such smoothing must be further analyzed in our framework, since it reduces noise but effectively also smooths sharp contrasts and levels of detail. Also, though there are proven advantages of tracking against the growing full surface model with respect to frame-to-frame tracking, there is still likely to be drift over long sequences, which will ultimately affect accuracy. Our intended use of ground-truth information effectively eliminates drift, aiding in more-accurate pose estimation and hence scene reconstruction. Furthermore, our framework can fuse any source of positional information, such as embedded internal sensors, and by virtue potentially capture geometry that is not visible to the naked eye. In the next section, work on viewing such “hidden” geometry is discussed.

C. Internal Imaging

Research in internal imaging has been existent for a long time and has resulted in tremendous advances in both medicine and security. Breakthroughs in these domains have largely been a result of physics and biology research. For example, magnets are used to image the gastrointestinal track, radio waves produced by excited hydrogen atoms are used to image the brain (Magnetic Resonance Imaging), and x-rays (Computerized Axial Tomography) and small amounts of radioactive substances (Positron Emission Tomography) are used for tomography. All of these procedures were revolutionary and are now routine. Similar techniques and technologies have recently been used in tandem with computer vision to enhance security. For example, Taddei and Sequeira demonstrated that x-ray tomography equipment could be calibrated using Automatic Pose Estimation (APE) on the silhouettes of shapes [13].

We are also motivated by structural health monitoring, which is concerned with using embedded sensor networks to evaluate structures such as buildings and bridges and provide feedback on cracks, torsion, and other instabilities. For a great overview see Tignola et al. [14]. We believe the sensors and technology used in structural health monitoring will eventually be miniaturized and can thus be expanded upon to be utilized for geometric reconstruction on much smaller objects than bridges. This has motivated us to begin our preliminary exploration of using such internal sensor networks and fusing them with imagery to improve current reconstruction approaches.

III. Sensor-Aided Reconstruction

In order to provide a valid comparison point for our approach, the first thing we did was perform our own re-constructions using structure-from-motion, space carving, and KinectFusion using a new dataset. For these reconstructions, we used an additively manufactured chess piece and Utah teapot, both in red polylactic acid (PLA) plastic via the Makerbot Replicator 2®. Our motivation in creating this new dataset using 3D printed objects was *a)* we have ground-truth knowledge about the correct geometry *b)* we intend to show the potential benefits of embedding sensors as part of the manufacturing process and *c)* it allows for the object to be materialized by anyone who wishes to try their own physical approach (for example, KinectFusion).

Our results, shown in Figure 2, demonstrate the various problems with these standard approaches for the objects in (a) and (e). Structure-from-motion results, (b) and (f), are meshed reconstructions retrieved from running Visual SfM [15][16] and using Patch-based Multi-view Stereo (PMVS) [17] to densify. While this approach does a decent job of accurately capturing some important details like the crown on the queen and the handle and spout of the teapot, it is clear that the reconstruction is full of holes and not dense enough. It is important to note that the lack of texture and color variance is one of the major problems for structure-from-motion since it largely depends on the presence of lots of unique features for tracking. Another common approach that doesn't depend on texture or color is space carving [18]. In images (c) and (g), you can see that although it does a great job of creating a dense, water-tight, reconstruction by virtue of the approach, it is not accurate enough to capture the sharp tips of the crown and none of the spout or handle of the teapot. Similarly, KinectFusion [1], (d) and (h), creates wonderfully dense objects but fails to capture small details due to the smallness of the objects, inherent noise and hardware limitations. It is also important to note that although these methods will yield better results for larger objects, the results are only aesthetically pleasing and not actually precise, hence why small features will be missed.

In light of these results, we developed an alternative reconstruction pipeline which couples positional information with structure-from-motion. In general, accurate structure-from-motion based reconstruction typically relies on accurate *feature tracking* [4]. A feature track is a set of pixel positions representing a scene point tracked over a set of images. Given a 3D position computed from multi-view stereo, its *reprojection error* with respect to its corresponding feature track is the only valid metric to assess error, in the absence of ground truth. Highly inaccurate individual track positions adversely affect subsequent camera pose estimation and structure computation, as well as bundle adjustment. Such inaccuracies can be improved upon by including external sensor information, such as positional information, into solving for scene reconstruction. The advantage of counting with embedded positional information inside an object is that it avoids having to compute accurate feature tracks in order to perform camera parameter and structure computation.

A diagram of our pipeline is shown in Figure 3. The process begins by collecting both the positional sensor data and image data. Provided with a mechanism for locating the positional sensor in each image, the accurate position information is used to perform camera pose estimation. This leads to accurate camera rotation and translation measurements and is void of the inaccuracies present when using feature tracks to estimate camera pose. Feature

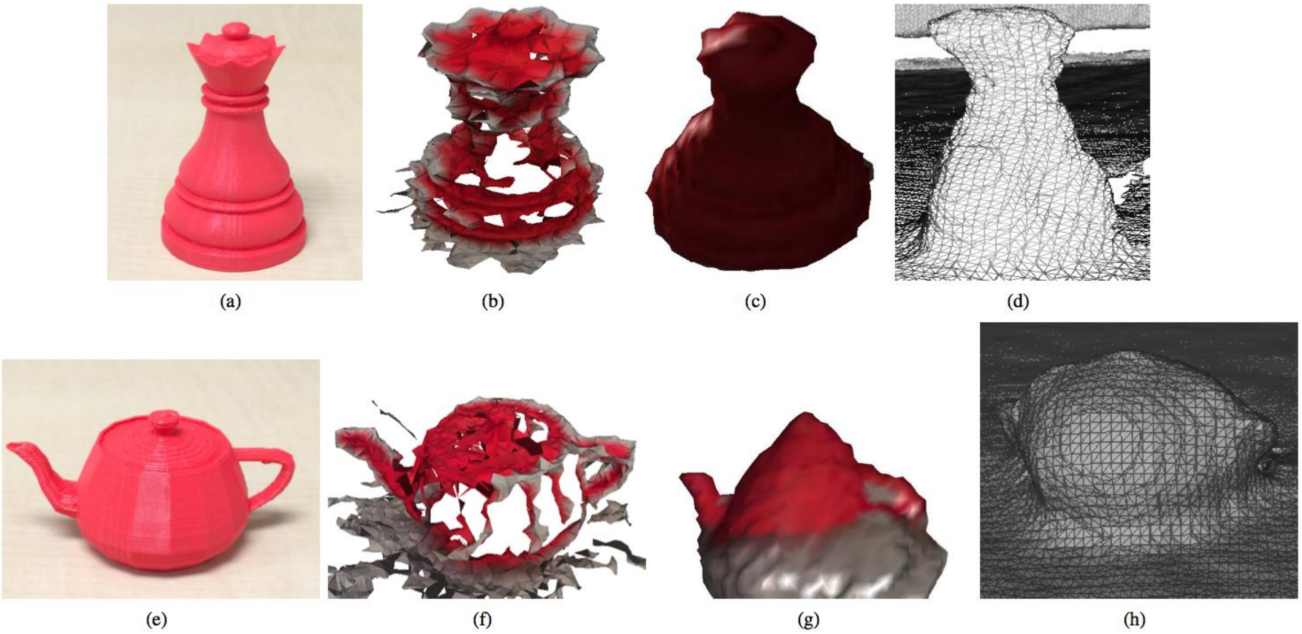


Figure 2. Input (a), structure-from-motion (b), space carving (c), and KinectFusion (d) reconstructions for a 3D-printed red chess piece. The same is shown for the "Utah teapot" in (e) - (h).

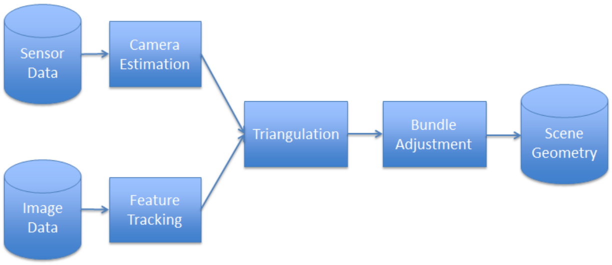


Figure 3. Our proposed reconstruction pipeline that utilizes sensors.

tracking is performed on the image data and is combined with the accurate camera data to perform triangulation of the scenes 3D structure. Errors in the feature tracking stage are manifested as inaccurate scene points and bundle adjustment is used to optimize the reprojection error of the given structure and camera parameters. After bundle adjustment has been performed, the scene geometry can be stored and manipulated using standard modeling techniques.

While our pipeline is only a work-in-progress, initial results show the positive effect of embedded positional sensors on reconstruction. We successfully performed a simulated reconstruction using synthetic data from the chess piece's geometrical definition (a .obj file) in order to sanity-check the camera estimation portion of our pipeline. To simulate surface-level embedded sensors we chose 185 random vertices from the definition file, whose locations appear in Figure 4a. By creating 10 randomly placed synthetic cameras (not pictured) and reprojecting the sensor locations into the synthetic image plane of each camera, we created feature tracks for each sensor. Using these feature tracks and the corresponding ground-truth locations of the sensors, we performed camera pose estimation using the Efficient Perspective-n-point (EPNP) algorithm [19]. Using feature tracks for all 18504 ground-truth vertices and using our computed cameras to triangulate we were able to achieve a reconstruction with essentially zero reprojection error (see Figure 4b). While this result is expected given that we have perfect feature tracks, we have shown that the camera pose estimation section of our pipeline has been implemented correctly and embedded sensors can be used for nearly perfect camera pose estimation. To complete our work, we would use additional feature tracks derived from SIFT-analyzed photography of an object with real surface-level embedded sensors, we discuss how to do so and the implications in the next section.

IV. Conclusion and Future Work

This paper presented the general idea of using sensor fusion as a strong tool for improving accuracy in computer vision structure-from-motion with the end goal of enabling high accuracy applications. Concrete results were shown for synthetic data, where a simulated object with surface-level position sensors was used to very accurately estimate the set of cameras viewing the object. Given these initial results, we believe the future is bright with regards to fusing sensor measurements for improved multi-view reconstruction, which is the focus of our ongoing work.

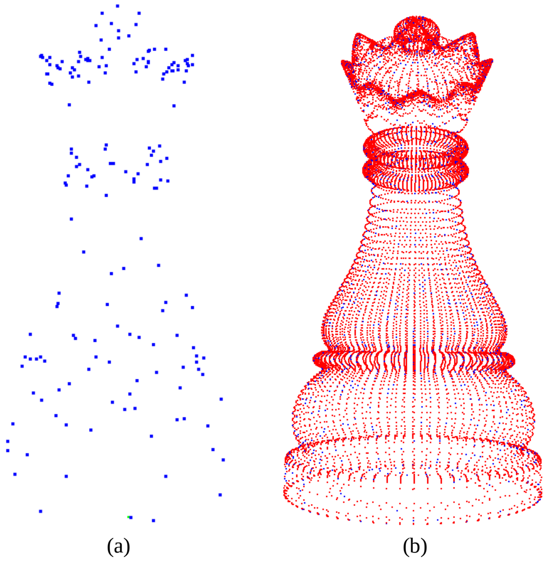


Figure 4. (a) Simulated surface-level embedded sensor locations. (b) A reconstruction using synthetically perfect feature tracks but computed cameras confirms nearly perfect camera pose estimation.

We have identified a number of uses in potential applications. One is the additive manufacturing process. By embedding positional sensors as part of the 3D-printing process, a whole host of opportunities open up. First, you can monitor and analyze the object during the printing process and verify key geometric qualities, such as distances or angles. Secondly, if the sensors are miniaturized to a sufficient degree and placed very densely, it becomes unnecessary to even use structure-from-motion or other techniques since a meshed point cloud of sensor locations can be used as a reconstruction by itself (see Figure 4b). Third, a designer could manipulate the printed object with real world tools, such as chisels and saws, and be able to ”scan” the object back into virtual space. A similar process already occurs in structural health monitoring where sensors are mixed with concrete; it is our belief that is only a matter of time before the technology is miniaturized enough for small scale objects and additive manufacturing.

Furthermore, at that miniaturized scale, we can expand the concept of structural health monitoring to medical applications and devices. By embedding sensors in artificial human parts, such as hearts and prosthetics, we enable the medical community to non-invasively monitor any defects that may occur by periodically reconstructing the object and analyzing it.

We strongly believe that miniaturized positional sensors are achievable with some combination of modern technologies such as ultrasound, magnets, and piezoelectrics. Simpler, surface-level sensors requiring manual effort could be created with highly reflective targets or glow-in-the-dark plastic/stickers for easy 2D localization by hand or automated process. Our future work will focus on using prototyped positional sensors to proof-of-concept our approach and its revolutionary applications.

ACKNOWLEDGMENT

This work is supported by the Laboratory Directed Research and Development program at Sandia National Laboratories, a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04- 94AL85000. We give special thanks to our colleagues at the Institute for Data Analysis and Visualization for the useful discussions and support.

REFERENCES

[1] S. Izadi et al. "Kinectfusion: Real-time 3d reconstruction and interaction using a moving depth camera," in Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology, ser. UIST '11. New York, NY, USA: ACM, 2011, pp. 559–568. [Online]. Available: http://doi.acm.org/10.1145/2047196.2047270 [retrieved: April, 2014]

[2] S. M. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski, "A Comparison and Evaluation of Multi-View Stereo Reconstruction Algorithms," in CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC, USA: IEEE Computer Society, 2006, pp. 519–528.

[3] D. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal On Computer Vision, vol. 60, no. 2, pp. 91–110, 2004.

[4] R. I. Hartley and A. Zisserman, Multiple View Geometry in Computer Vision, 2nd ed. Cambridge University Press, 2004.

[5] M. Lourakis and A. Argyros, "The design and implementation of a generic sparse bundle adjustment software package based on the Levenberg-Marquardt algorithm," Institute of Computer Science - FORTH, Heraklion, Crete, Greece, Tech. Rep. 340, August 2000.

[6] Y. Chen and G. Medioni, "Object modelling by registration of multiple range images," Image Vision Comput., vol. 10, no. 3, pp. 145–155, Apr. 1992. [Online]. Available: http://dx.doi.org/10.1016/ 0262- 8856(92)90066- C

[7] J. Tong, J. Zhou, L. Liu, Z. Pan, and H. Yan, "Scanning 3d full human bodies using kinects," Visualization and Computer Graphics, IEEE Transactions on, vol. 18, no. 4, pp. 643–650, 2012.

[8] A. Majdi, M. C. Bakkay, and E. Zagrouba, "3d modeling of indoor environments using kinect sensor," in Image Information Processing (ICIIP), 2013 IEEE Second International Conference on, 2013, pp. 67–72

[9] R. A. Newcombe et al. KinectFusion: Real-Time Dense Surface Mapping and Tracking, in IEEE ISMAR, IEEE, October 2011.

[10] G. Klein and D. W. Murray. Parallel tracking and mapping for small AR workspaces. In Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR), 2007.

[11] R. A. Newcombe and A. J. Davison. Live dense reconstruction with a single moving camera. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2010.

[12] J. Stuehmer, S. Gumhold, and D. Cremers. Real-time dense geometry from a handheld camera. In Proceedings of the DAGM Symposium on Pattern Recognition, 2010.

[13] P. Taddei and V. Sequeira, "X-ray and 3d data fusion for 3d reconstruction of closed receptacle contents," in 3DV-Conference, 2013 International Conference on, 2013, pp. 231–238.

[14] D. Tignola, S. Vito, G. Fattoruso, F. D'Aversa, and G. Francia, "A wireless sensor network architecture for structural health monitoring," in Sensors and Microsystems, ser. Lecture Notes in Electrical Engineering, C. Di Natale, V. Ferrari, A. Ponzoni, G. Sberveglieri, and M. Ferrari, Eds. Springer International Publishing, 2014, vol. 268, pp. 397–400. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-00684-0 76 [retrieved: April, 2014]

[15] C. Wu, S. Agarwal, B. Curless, and S. Seitz, "Multicore bundle adjustment," in Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on, 2011, pp. 3057–3064.

[16] C. Wu, "Towards linear-time incremental structure from motion," in 3DV-Conference, 2013 International Conference on, 2013, pp. 127–134.

[17] Y. Furukawa and J. Ponce, "Accurate, dense, and robust multi-view stereopsis," in IEEE Conference on Computer Vision and Pattern Recognition, June 2007, pp. 1–8.

[18] K. Kutulakos and S. Seitz, "A theory of shape by space carving," International Journal of Computer Vision, vol. 38, no. 3, pp. 199-218, 2000. [Online]. Available: http://dx.doi.org/10.1023/A%3A1008191222954 [retrieved: April, 2014]

[19] Vincent Lepetit, Francesc Moreno-Noguer, and Pascal Fua. 2009. EPnP: An Accurate O(n) Solution to the PnP Problem. Int. J. Comput. Vision 81, 2 (February 2009), 155-166. DOI=10.1007/s11263-008-0152-6 http://dx.doi.org/10.1007/s11263-008-0152-6

[20] Besl, P.J.; McKay, Neil D., "A method for registration of 3-D shapes," Pattern Analysis and Machine Intelligence, IEEE Transactions on , vol.14, no.2, pp.239,256, Feb 1992 doi: 10.1109/34.121791 [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=121791&isnumber=3469 [retrieved: April, 2014]

[21] Newcombe, Richard A.; Lovegrove, S.J.; Davison, A.J., "DTAM: Dense tracking and mapping in real-time," Computer Vision (ICCV), 2011 IEEE International Conference on , vol., no., pp.2320,2327, 6-13 Nov. 2011 doi: 10.1109/ICCV.2011.6126513 [Online]. Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6126513&isnumber=6126217 [retrieved: April, 2014]

PRIMA - Towards an Automatic Review/Paper Matching Score Calculation

Christian Caldera, René Berndt, Eva Eggeling
Fraunhofer Austria Research GmbH
Email: {christian.caldera, rene.berndt, eva.eggeling}
@fraunhofer.at

Martin Schröttner
Institute of Computer Graphics and Knowledge Visualization
University of Technology, Graz, Austria
Email: martin.schroettner@cgv.tugraz.at

Dieter W. Fellner
Institute of ComputerGraphics and KnowledgeVisualization (CGV), TU Graz, Austria
GRIS, TU Darmstadt & Fraunhofer IGD, Darmstadt, Germany
Email: d.fellner@igd.fraunhofer.de

***Abstract*—Programme chairs of scientific conferences face a tremendous time pressure. One of the most time-consuming steps during the conference workflow is assigning members of the international programme committee (IPC) to the received submissions. Finding the best-suited persons for reviewing strongly depends on how the paper matches the expertise of each IPC member. While various approaches like ”bidding” or ”topic matching” exist in order to make the knowledge of these expertises explicit, these approaches allocate a considerable amount of resources on the IPC member side. This paper introduces the *Paper Rating and IPC Matching* Tool (PRIMA), which reduces the workload for both - IPC members and chairs - to support and improve the assignment process.**

Keywords-Conferences, International Program Committee, Submissions, Paper, Assignment, Matching, TF-IDF, Information Retrieval.

I. Introduction

Conferences and journals play an important role in the scientific world. Both are important channels for exchange of information between researchers. The publication list of a researcher defines his/her standing within the scientific community. In order to ensure quality standards for these publications, submitted work undergo the so called peer review process. This process is used to maintain standards, improve performance and provide credibility [1]. Today almost every conference or journal uses a electronic conference management system in order to organize this process.

In-a-nutshell the peer review process for a conference works as follows: First, authors upload their paper to the electronic submission system. After the deadline the submitted papers are distributed to the reviewers. The conference reviewers are usually members of the International Programme Committee (IPC) and - depending on the size of the conference - a pool of external experts. Each reviewer receives a certain amount of submitted papers depending on his/her expertise. Assigning the submitted papers to the IPC members is a crucial task in the peer review process because these reviews decide if the paper is accepted or not. In case of acceptance the author is allowed to upload a camera-ready-copy version of the paper, which is then published in the proceeding of the conference.

This is the essence of the peer review process. Within the peer review process there exist variations, which mostly differ in what information is revealed to whom. The most commonly used are the single and double blinded peer reviewing process:

- In the single blinded peer review the identity of the reviewer is unknown to the user. But, the reviewer knows the identity of the author. In this setting, the reviewer can give a critical review without the fear that the person itself will be targeted by the author.
- In the double blinded peer review, the identity of the reviewer and author is unknown to each other. This process guarantees the same chances for unknown and famous scientist and universities by removing the name on the submissions.

There are further versions of peer reviewing like open peer reviewing or additions like post-publication peer reviewing, but they are rarely been applied [2][3].

The crucial step for the quality of the peer review process is to find the best suited reviewer for each of the submitted papers. This person must fulfill the following two conditions:

- He should be an expert on the topic of the paper in order to give a qualified judgment on novelty, contribution and other aspects of the work presented.
- He must not be in any kind related to the author to guarantee a neutral statement about this paper, which means that there is no conflict.

After the conflicts of the reviewer are identified, he needs to be assigned to one or multiple papers. But before they can be assigned, an indicator is needed to measure which reviewer suits best to which paper. Most systems use a so called bidding process. In this case, the IPC member can indicate what papers he wants to review and in which areas he considers himself as an expert. This process is tedious for IPC members of conferences with hundreds of submissions. To address this problem, this paper presents an automatic approach by using the TF/IDF algorithm for matching the submitted papers with existing publications of the IPC member.

The next section will give an overview of other systems and techniques used in this domain. Section III and IV will introduce the TF/IDF and our implementation of PRIMA. The last two sections will address the results and the future work on PRIMA.

II. Related Work

There has already been some research on how to create a good matching between a reviewer and submission. Charlin et al. created a framework based on a machine learning techniques [4]. Dumais et al. examined Latent Semantic Indexing methods [5] for assigning reviewers to submissions. Hettich et al. and Basu et al. extracted with TF-IDF important words in submissions and mined the web for possible reviewers based on the extracted TF-IDF terms [6], [7]. In the context of submission paper to IPC matching further problems arise when there is a given amount of IPC members:

A. *Conflict detection*

Current systems use different approaches for conflict detection. For example, in EasyChair, one of the largest conference management systems [8], the IPC member manually specifies for which papers he has a conflict with the author [9].

Confious [10] uses an automatic approach in order to detect conflicts by comparing email suffixes or affiliation data. The problem there is that people do have more than one email address and they often do not use their institutional email address but rather an address from a large email service (e.g., Yahoo, GMail, GMX, etc.).

A more robust approach in finding these conflicts is implemented in SRMv2 [11] by queering the Digital Bibliography & Library Project (DBLP) [12] and checking if the IPC member and the author have a co-authorship. If there is a co-authorship found on the DBLP it indicates also a conflict of interests for further submissions [13].

B. *Reviewer suitability*

The reviewer of a paper must be an expert in the area of the paper. For this reason, the review assignment can not be done on a random basis. So the conference management system needs a method to rate how suitable a reviewer is for a submission. Many of the current systems use some kind of bidding mechanism to generate these values. These bidding systems can be separated in two classes:

- An IPC member manually bids on the areas of expertise. During the submission phase an author can classify his paper according to a predefined topic list. These topics can be special areas defined for a conference or a general classification scheme for example the ACM classification [14]. The IPC member receives the same list in order to define his own preferences in what fields he considers himself as an expert. An IPC member who is an expert in an area is a possible candidate for reviewing papers of this specific area.
- The IPC member manually bids directly on the papers. Based on the title and abstract of a submission the members can decide if they are qualified to review it or not.

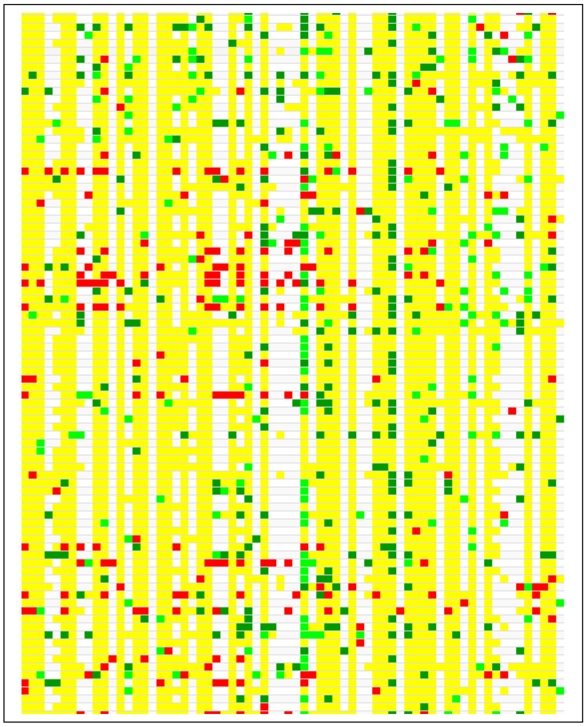


Figure 1: This figure shows a screen shot of the global bidding matrix. There it can be seen that the default value and the empty values take up most of the bidding values.

Larger conferences with hundreds of submissions sometimes combine these two options. As it is a considerable effort for an IPC member to read over hundreds of titles and abstracts, some systems let the users first fill out the topic list. Based on these data the systems generate a pre-ordered list of submissions for each IPC member. After that, the IPC members can read the title and abstracts of these submission, which fit best to their expertise profile. The resulting ratings can be used in the assignment process. This system however has two major drawbacks:

- Rodriguez et al. show in their paper ”Mapping the Bid Behavior of Conference Referees” [15] that human-driven referee bidding may not be the best solution for conference bidding due to referee fatigue. After reading several titles and abstracts an IPC member can have another decision basis. Furthermore, this bidding technique is susceptible to sloppy biddings due to curiosity or to unclear title and abstract of a paper. Using this bidding method, an IPC mehis area of expertise.
- The second issue about this system is that it doesn’t scale very well. An IPC member might be fine with reading some titles and abstracts. But if a conference has several hundreds submissions the effort for an IPC members is too large. Furthermore, it is not reasonable for all members to read all abstracts and have the same objectivity towards the last papers compared to the first. In addtion, if a reviewer reads only the papers in his expert area some papers which would fit to his expertise might be unnoticed.

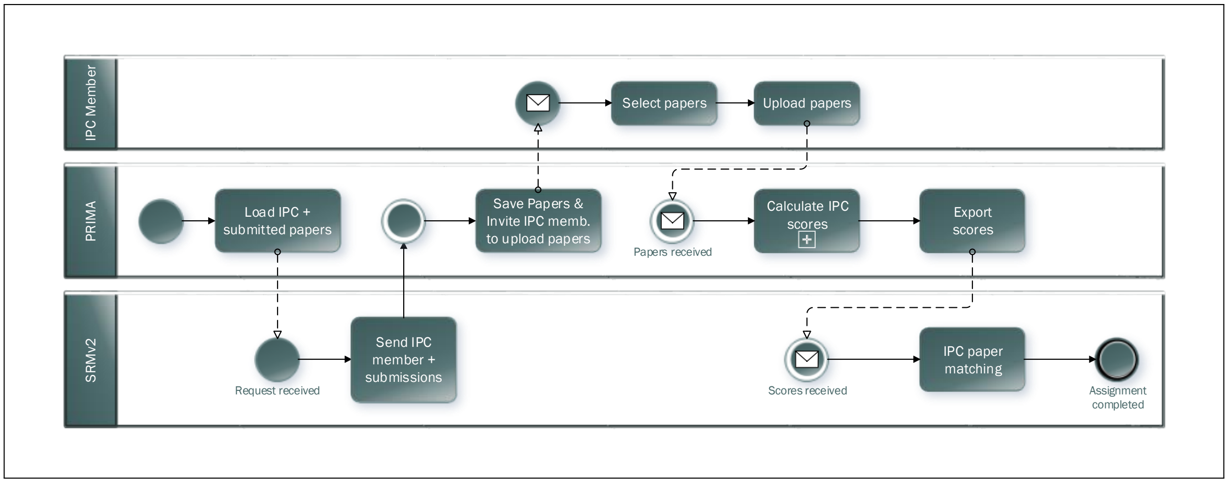


Figure 2: This figure shows the PRIMA workflow: how PRIMA receives the data and invites the IPC members to upload their papers. When all papers are received the calculation can start. After that the final scores can be exported again.

Figure 1 shows the bidding matrix for an exemplary conference. In this examples the IPC members (columns) specified for all papers (rows), whether they *want to review* (dark-green), *could review* (light-green), *have a conflict* (red) or are *not competent* (yellow). A large portion of the bidding matrix is either not filled up (white cells) or marked as *not competent* (yellow cells).

Our approach towards these problems is to automate the reviewer suitability rating by using the *Term Frequency Inverse Document Frequency* (TF/IDF) in order to categorize the submitted papers with respct to existing publications of the IPC members. These generated values can be used to refine and improve the values from the manual bidding process or even make the manual process obsolete. The huge advantage of this approach is, that it is possible to create a better IPC to paper distribution instead of a distribution of more or less randomly assigned reviewers.

III. TF/IDF

This section gives a small introduction to the mathematical basis of the TF/IDF. The term TF/IDF stands for *Term Frequency Inverse Document Frequency*. The TF/IDF algorithm can be separated into two parts. The first part is the *Term Frequency* part. As the name suggests it uses the frequency of terms in a document to classify the document. The second part of the algorithm is the *Inverse Document Frequency*. This means that the terms are weighted according to the occurrence in several documents. That is the more a term is used in different documents the less information it provides for classifying a document [16].

This paper will give an overview how the algorithm works. The algorithm itself is already a quite understood and researched topic in different areas like text categorization, text analysis, mining and information retrieval techniques.

In the term frequency calculation (see (1)) every term t in the document d is counted. For weighting the different terms in the document the logarithm is calculated. This is done because a term which occurs 10 times more than another term is not 10 times more meaningful.

$$tf(t,d) = log(1 + f(t,d)) \tag{1}$$

The inverse document frequency (see (2)) counts the occurrences of a term across all documents in a given document corpus. This is done by taking the logarithm of the quotient between the total number of documents $|D|$ and the amount of documents d containing the term t. Imagine a term which occurs in every document. This term is not useful for categorizing so it has to be penalized for being not important in the current global text corpus. Terms which occur in fewer documents receive a higher value with this formula.

$$idf(t,D) = log\frac{|D|}{|\{d \in D : t \in d\}|} \tag{2}$$

By multiplying the term frequency with the inverse document frequency the TF/IDF is received (see (3)). This value classifies a term in a document and its classification significance across all documents [17].

$$tfidf(t,d,D) = tf(t,d) \ \times \ idf(t,D) \tag{3}$$

All TF/IDF values of a document form a vector which classifies the document. By calculating the cosine similarity (see (4)) between two documents it is then possible to extract a similarity value [18]. By calculating the similarity between all submitted papers and previous papers of the IPC members we want to extract a matching value which enables matching submissions and IPC members together.

$$\cos(a,b) = \frac{a \cdot b}{\parallel a \parallel \parallel b \parallel} \tag{4}$$

There are many abbreviations in the TF/IDF algorithm, which offer different advantages and disadvantages. In the current version of our implementation the above described TF/IDF algorithm is used. Further research will show if another version or combination of algorithm yield to better results.

IV. PRIMA - Paper Rating and Ipc MAtching Tool

This section describes the prototype of the *Paper Rating and IPC Matching Tool* (PRIMA), which is a standalone extension to the SRMv2 conference management system. The workflow of the automatic score generation with PRIMA is shown in Figure 2.

In the first step, the PRIMA tool is initalized with the required data for the IF/IDF calculation: the submitted paper along with their metadata and information about the IPC members of the event. PRIMA uses the API of the SRMv2 framework [13] in order to fetch the required information. After the initalization, the IPC members are invited to upload their publications which fit best to the scope of the conference. The more papers a user uploads into the system the better the algorithm can find different matchings to the submissions of the conference.

One critical issue we found during the tests was that some IPC members received an overall good score on every paper. During the investigations we found out that some of the uploaded data were conference proceedings. From this files only the paper of the person was extracted and analysed as the whole report distorted the expertise of the user.

Then for all submitted papers of the conference and all uploaded publications of the IPC members, the paper scores are calculated. Then, these scores are transmitted into SRMv2 in order to support the pre-ordering for the bidding process and to support the assignment process.

Before the calculation itself starts some preprocessing steps are necessary to improve the TF/IDF result:

- For all uploaded publications, the raw text is extracted from the Portable Document Format (PDF) documents. This extracted text contains a large number of unnecessary information, which do not have an impact on the paper classification, for example numbers, special characters, code, urls, email addresses, punctuation, authors, addresses, IDs, etc. Future work on TF/IDF concentrates how to separate the text which is useful for the TF/IDF score generation from the overhead part which interferes with the generation [19].
- In the next step, stop words are removed. Stop words are words which occur often in a text but do not add any informational value to the text. Some examples of this stop words are: and, or, the, an, important, however, just and so on. All these words are necessary for the creation of sentences. But two texts do not relate strongly to each other just because they have a lot of “and” together [20].
- In the last step before the TF/IDF is applied all words have to be stemmed. Stemming reduces words to their common root. For example ”overview” and ”overviews” are not the same words in a computational matching, so the word overviews is reduced to overview. These two words will then match in the algorithm [21].
- The TF/IDF algorithm counts the words, normalizes and then they are weighted according to the occurrences in the other documents see Section III.
- After the TF/IDF calculation has been completed, each submission is compared against all papers of the IPC members with the cosine similarity.

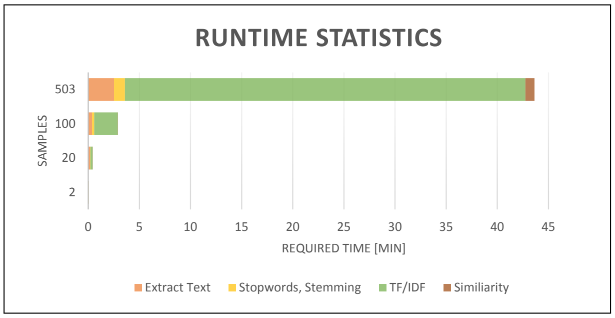


Figure 3: Runtime Statistic. This figure shows the amount of time each of the tasks take. It can be seen that the algorithm has an exponential growth.

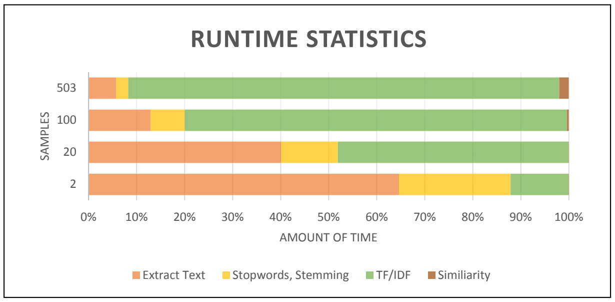
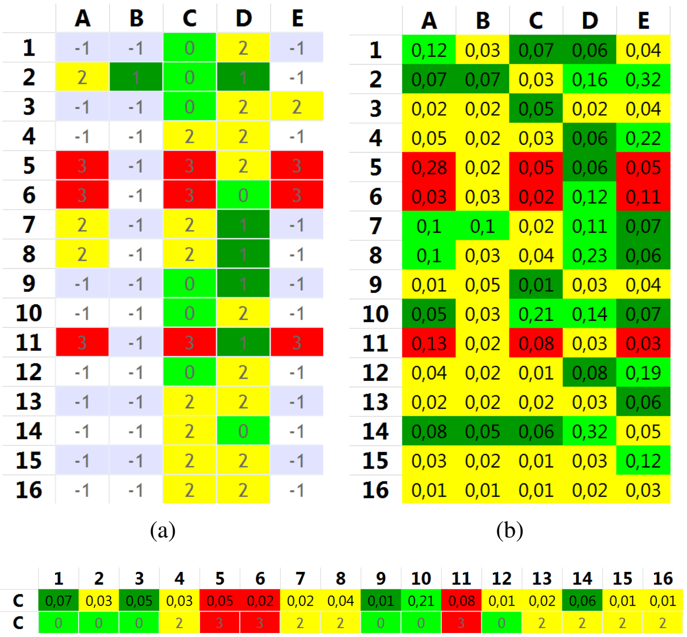


Figure 4: Runtime Statistics. This figure shows the amount of time each of the tasks take, split by the tasks and scaled to 100%.

If an IPC member has provided multiple publications, all of them are checked against a single submission paper. Currently, the average of the best five papers is saved. This is done to prevent statistical outliers. Furthermore, not all papers are taken into consideration as a person might upload a lot of papers belonging to different areas. In this case, every area on its own would have a lower average which falsifies the expertise area of a person. Further research will show if other values or a special algorithm should be used for creating a stronger statement about a submission and an IPC member.

V. Results

For testing the data the Eurographics 2014 was chosen. The papers, the submissions and the reviewers are anonymized and randomly reordered. Here are some statistics about the conference: There were about 70 programme committee members and about 290 submissions. Every IPC member entered their conflicts, defined areas of expertise to create a pre-filtering for the submissions and finally bidded on the paper. This final bidding matrix has 290 x 70 = 20300 entries (see Figure 1). For testing in PRIMA, about 300 papers of these IPC members were uploaded and together with the 290 submissions analyses through the TF/IDF algorithm.



	A	B	C	D	E
1	-1	-1	0	2	-1
2	2	1	0	1	-1
3	-1	-1	0	2	2
4	-1	-1	2	2	-1
5	3	-1	3	2	3
6	3	-1	3	0	3
7	2	-1	2	1	-1
8	2	-1	2	1	-1
9	-1	-1	0	1	-1
10	-1	-1	0	2	-1
11	3	-1	3	1	3
12	-1	-1	0	2	-1
13	-1	-1	2	2	-1
14	-1	-1	2	0	-1
15	-1	-1	2	2	-1
16	-1	-1	2	2	-1

(a)

	A	B	C	D	E
1	0,12	0,03	0,07	0,06	0,04
2	0,07	0,07	0,03	0,16	0,32
3	0,02	0,02	0,05	0,02	0,04
4	0,05	0,02	0,03	0,06	0,22
5	0,28	0,02	0,05	0,06	0,05
6	0,03	0,03	0,02	0,12	0,11
7	0,1	0,1	0,02	0,11	0,07
8	0,1	0,03	0,04	0,23	0,06
9	0,01	0,05	0,01	0,03	0,04
10	0,05	0,03	0,21	0,14	0,07
11	0,13	0,02	0,08	0,03	0,03
12	0,04	0,02	0,01	0,08	0,19
13	0,02	0,02	0,02	0,03	0,06
14	0,08	0,05	0,06	0,32	0,05
15	0,03	0,02	0,01	0,03	0,12
16	0,01	0,01	0,01	0,02	0,03

(b)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
C	0,07	0,03	0,05	0,03	0,05	0,02	0,02	0,04	0,01	0,21	0,08	0,01	0,02	0,06	0,01	0,01
C	0	0	0	2	3	3	2	2	0	0	3	0	2	2	2	2

(c)

Figure 5: Figure (a) shows a small excerpt of the bidding matrix. Most reviewers set the values to the default not competent or did not submit any values at all. Figure (b) shows an excerpt of the PRIMA matrix with the same color encoding like the bidding matrix and thresholds at 0.05 and 0.1. Figure (c) shows the transposed bidding and calculated matrix of reviewer C for easier comparison.

Figure 5a shows a small excerpt of the bidding matrix. The rows represent five reviewers (a to e), the columns represent 16 submissions (1 - 16). The colors are encoded in the following way: Green means the IPC member submitted that he is able to review the paper. Whereas 0 (light green) means he wants to review the paper and 1 (dark green) means he could review the paper. 2 (yellow) indicates that the reviewer said he is not competent enough to review this paper. From -1 (white areas) we do not have any data as the IPC member did not bid on this paper. The red spots mark a conflict of the IPC member with the author of the submitted paper.

Figure 5b shows the same excerpt for the TF/IDF algorithm. The algorithm outputs a value between 0 and 1. Where 0 means no word overlap in both documents and 1 means every word in both papers appear at the same amount. Based on the global output we inked the output of the algorithm to give a similar appearance like the bidding matrix. The threshold of the values are 0.05 and 0.1. Everything below 0.5 is colored in yellow meaning a low correlation. Between 0.05 and 0.1 a medium correlation exists, which are marked in dark green. The high correlation (> 0.1) are colored in light green. Additional the conflicts of the bidding are included in the results. To better compare these two tables the third figure shows the bidding matrix of the reviewer C to the calculated values.

A first observation is, that the left bidding figure shows that the provided data of the IPC is rather incomplete. This might happen because an IPC only checked the papers in his own area of expertise or because of a lack of time he was not able to read all 290 submission abstracts.

Another important observation is that some good matchings are conflicts (see at cell C11 of the left figure). This shows that the approach itself is heading in the right direction as it can be expected that a person who is an expert in an area also might have a project cooperation other experts in this field and therefore has a conflict of interests with this person.

Furthermore, it can be observed that most of the bidding match with the found generated values of PRIMA C1, C3, C9, C10 (Figure 5c). In addition, also the not competent column matches with the biddings C4, C7, C8, C13, C16. In this case, it should be said that the uploaded data of each person were taken only from previous Eurographics events and that the amount of uploaded data also differs. For example IPC member D has 18 uploaded papers and person B only five. For this reason person D is much better classified by the TF/IDF and therefore has a better matching than person B.

Strong differences between the bidding and the calculated classification, e.g., for person C the cells C2, C12, and C14, can have multiple reasons.

According to the TF/IDF the IPC member would be well suited as a reviewer, but he considered himself as not competent. This can have different reasons:

- The TF/IDF has analyzed an older paper of the person, but the expertise focus of the person has changed.
- The title and abstract from the bidding might have been misleading.
- The submission was overlooked by the IPC member and this submissions stayed on the default value which is *not competent*.

The first item will be addressed in further research in order to analyze if penalty value for older paper will improve the results.

But also cases where the rating from the TF/IDF shows a low score, but the persons claimed that he *wants to review* occur, for example in C2 and C12:

- Most likely the system does not have a current paper of the IPC member on this topic.
- The reviewer is interested in a paper and ”wants to review” it, but does not have the necessary knowledge to review it.

As stated before a large portion of the bidding matrix is not filled up (see Figure 1). One huge advantage of the process is when there is no value on the bidding matrix but the calculation found excellent matches on the algorithm. There it is possible to create a better reviewer-to-paper assignment instead of randomly distributing the submitted papers to the reviewers. For example in submission 4 the best matches are person D and E, for submitted paper 13 person E would be a good choice.

Figure 4 and Figure 3 show the runtime statistics of the PRIMA tool split into the four steps *extract text*, *stopwords,* *stemming*, *TF/IDF*, and *similiarity*. It can be seen that for a small number of papers the extraction of the text and the stopwords removal and stemming takes up most of the time. If the number of papers increases, the more time the TF/IDF algorithm itself takes. The text extraction and stopwords removal and stemming can be precalculated and stored. However, this step takes less than 10% of the time during the full calculation using more than 500 papers. The TF/IDF itself cannot be precalculated as every further submission changes the weighting of each word in the calculation process. So it has to be calculated when all papers of the IPC members are available. The algorithm on 503 papers takes up about 45 minutes, Where 100 papers only take about 3 minutes and 2 and 20 papers are calculated in seconds.

VI. Conclusion & Future Work

In this paper, we presented the PRIMA tool, which automatically calculates a ranking between submitted papers and the available reviewers (IPC members). By using the TF/IDF for categorizing the submitted papers along the reviewers expertise, the workload of the reviewers and the chairs is reduced dramatically. TF/IDF itself is already a well researched topic in text categorization and information retrieval techniques.

One large problem in comparing various tools and their performance is the lack of a standardized benchmark for this task. All work so far, has used real data for evaluating and testing. Since they contain sensitive information, these datasets cannot become publicly available, which makes a direct comparison impossible.

For the upcoming Europgrahics conference it is planned to evaluate the scores by presenting submissions to the authors in descending order. Then the IPC member can concentrate on the title/abstracts which fit best to the topics of his own publications. The values that the PRIMA tool generates can also be used as suggestions for the reviewer during the bidding process. This way the member can skim over the values and check if they fit.

Currently the selection and upload of publications is done manually by the reviewers. Using citation portals like DBLP [12], Citeseer [22] and other sources, the selection and retrieval of the full-text version (e.g., when available through the Open Access [23] initiative) can be automated as well.

Another important point which might be improved is the text extraction itself. At the moment, the whole paper is used for the TF/IDF. And although the numbers, special characters, URLs, stopwords, etc., are removed there are still words which slip through which should not be used for the analysis. For example words like the author, the institution, figure explanations, headings, formulas and so on.

References

[1] Academia Publishing, “What is Peer Review?” 2014, [retrieved: 03, 2014]. [Online]. Available: http://academiapublishing.org

[2] R. M. Blank, “The effects of double-blind versus single-blind reviewing: Experimental evidence from the american economic review,” American Economic Review, vol. 81, no. 5, December 1991, pp. 1041–67, [retrieved: 03, 2014]. [Online]. Available: http://ideas.repec.org/a/aea/aecrev/v81y1991i5p1041-67.html

[3] M. W. Consulting, “Peer review in scholarly journals: perspective of the scholarly community–an international study,” Author, Bristol, UK, 2008.

[4] L. Charlin and R. S. Zemel, “The toronto paper matching system: An automated paper-reviewer assignment system,” in ICML 2013 Workshop on Peer Reviewing and Publishing Models., Atlanta, Georgia, USA, Jun. 2013.

[5] S. T. Dumais and J. Nielsen, “Automating the assignment of submitted manuscripts to reviewers,” in Proceedings of the 15th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 1992, pp. 233–244.

[6] C. Basu, H. Hirsh, W. W. Cohen, and C. Nevill-manning, “Recommending papers by mining the web,” in Proceedings of the IJCAI99 Workshop on Learning about Users, 1999, pp. 1–11.

[7] S. Hettich and M. J. Pazzani, “Mining for proposal reviewers: lessons learned at the national science foundation,” in Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2006, pp. 862–871.

[8] L. Parra, S. Sendra, S. Ficarelli, and J. Lloret, “Comparison of online platforms for the review process of conference papers,” in CONTENT 2013, The Fifth International Conference on Creative Content Technologies, 2013, pp. 16–22.

[9] Cool Press Ltd, “EasyChair conference system,” 2013, [retrieved: 03, 2014]. [Online]. Available: http://www.easychair.org/

[10] M. Papagelis and D. Plexousakis, “Conf!ous - The Conference Nous,” 2013, [retrieved: 03, 2014]. [Online]. Available: http://www.confious.com/

[11] C. Caldera, “Srm 2.0,” 2013, [retrieved: 03, 2014]. [Online]. Available: https://srmv2.eg.org/COMFy

[12] M. Ley et al., “DBLP Computer Science Bibliography,” 2013, [retrieved: 03, 2014]. [Online]. Available: http://www.informatik.uni-trier.de/ ley/db/

[13] C. Caldera, R. Berndt, and D. W. Fellner, “Comfy - A Conference Management Framework,” Information Services and Use, vol. 33, no. 2, 2013, pp. 119–128, [retrieved: 03, 2014]. [Online]. Available: http://dx.doi.org/10.3233/ISU-130697

[14] The Association for Computing Machinery, Inc., “Association for Computing Machinery,” 2013, [retrieved: 03, 2014]. [Online]. Available: http://www.acm.org/about/class/class/2012

[15] M. A. Rodriguez, J. Bollen, and H. V. D. Sompel, “Mapping the bid behavior of conference referees,” Journal of Informetrics, vol. 1, 2007, pp. 06–0749.

[16] J. Ramos, “Using TF-IDF to Determine Word Relevance in Document Queries,” Department of Computer Science, Rutgers University, 23515 BPO Way, Piscataway, NJ, 08855e, Tech. Rep., 2003.

[17] C. D. Manning, P. Raghavan, and H. Schtze. Cambridge University Press, 2008, [retrieved: 03, 2014]. [Online]. Available: http://dx.doi.org/10.1017/CBO9780511809071.007

[18] A. Huang, “Similarity Measures for Text Document Clustering,” in New Zealand Computer Science Research Student Conference, J. Holland, A. Nicholas, and D. Brignoli, Eds., Apr. 2008, pp. 49–56. [Online]. Available: http://nzcsrsc08.canterbury.ac.nz/site/digital-proceedings

[19] E. Rahm and H. H. Do, “Data cleaning: Problems and current approaches,” IEEE Data Eng. Bull., vol. 23, no. 4, 2000, pp. 3–13.

[20] C. Silva and B. Ribeiro, “The importance of stop word removal on recall values in text categorization,” in Neural Networks, 2003. Proceedings of the International Joint Conference on, vol. 3, 2003, pp. 1661–1666 vol.3.

[21] M. Porter, “An algorithm for suffix stripping,” Program: electronic library and information systems, vol. 14, no. 3, 1980, pp. 130–137.

[22] The Pennsylvania State University, “CiteSeer,” 2014, [retrieved: 03, 2014]. [Online]. Available: http://citeseerx.ist.psu.edu/

[23] Georg-August-Universitt Gttingen Niederschsische Staats- und Universittsbibliothek Gttingen, “Open Access,” 2013, [retrieved: 03, 2014]. [Online]. Available: http://open-access.net