IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 6.252

Volume 2, Issue 1, November 2022

A Survey Study on Automatic Subtitle Synchronization and Positioning System for Deaf and Hearing Impaired People

Santosh S Kale[1], Shruti Dhanak[2], Paras Chavan[3], Jay Kakade[4], Prasad Humbe[5]

Guide, Department of Computer Engineering[1]

Students, Department of Computer Engineering[2,3,4,5]

NBN Sinhgad School of Engineering, Pune, Maharashtra, India

Abstract: *In this study, we provide a subtitle synchronisation and placement system intended to improve deaf and hearing-impaired individuals' access to multimedia content. The paper's main contributions are a novel synchronisation algorithm that can reliably align the closed caption with the audio transcript without any human involvement and a timestamp refinement technique that can modify the duration of the subtitle segments in accordance with audiovisual recommendations. Regardless of the kind of video, the experimental evaluation of the strategy on a sizable dataset of 30 films pulled from the French national television verifies the method with average accuracy scores above 90%. The success of our strategy is demonstrated by the subjective assessment of the suggested subtitle synchronization and location system, carried out with real hearing challenged persons.*

Keywords: Automatic Subtitle Synchronization.

REFERENCES

[1]. A. Katsamanis, M. P. Black, P. G. Georgiou, L. Goldstein, and S. Narayanan, ''SailAlign: Robust long speech-text alignment,'' in Proc. Workshop New Tools Methods Very-Large Scale Phonetics Res., Philadelphia, PA, USA, Jan. 2011, pp. 1–4.

[2]. X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, and J. Liang, ''EAST: An efficient and accurate scene text detector,'' in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Honolulu, HI, USA, Jul. 2017, pp. 2642–2651.

[3]. P. J. Moreno, C. Joerg, J.-M. Van Thong, and O. Glickman, ''A recursive algorithm for the forced alignment of very long audio segments,'' in Proc. Int. Conf. Spoken Lang. Process, Dec. 1998, pp. 2711–2714.

[4]. M. H. Davel, C. V. Heerden, N. Kleynhans, and E. Barnard, ''Efficient harvesting of Internet audio for resource-scarce ASR,'' in Proc. Interspeech, Aug. 2011, pp. 3154–3157.

[5]. N. Braunschweiler, M. J. F. Gales, and S. Buchholz, ''Lightly supervised recognition for automatic alignment of large coherent speech recordings,'' in Proc. Interspeech, Sep. 2010, pp. 2222–2225.

[6]. X. Anguera, J. Luque, and C. Gracia, ''Audio-to-text alignment for speech recognition with very limited resources,'' in Proc. Interspeech, Sep. 2014, pp. 1405–1409.

[7]. B. Axtell, C. Munteanu, C. Demmans Epp, Y. Aly, and F. Rudzicz, ''Touchsupported voice recording to facilitate forced alignment of text and speech in an E-Reading interface,'' in Proc. 23rd Int. Conf. Intell. User Interface, Mar. 2018, pp. 129–140.

[8]. I. Ahmed and S. K. Kopparapu, ''Technique for automatic sentence level alignment of long speech and transcripts,'' in Proc. Interspeech, Aug. 2013, pp. 1516–1519.

[9]. G. Hinton, L. Deng, D. Yu, G. E. Dahl, A. R. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, T. N. Sainath, and B. Kingsbury, ''Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups,'' IEEE Signal Process. Mag., vol. 29, no. 6, pp. 82–97, Nov. 2012, doi: 10.1109/MSP.2012.2205597.

[10]. N. T. Vu, F. Kraus, and T. Schultz, ''Rapid building of an ASR system for under-resourced languages based on multilingual unsupervised training,'' in Proc. Interspeech, Aug. 2011, pp. 1–4.

DOI: 10.48175/IJARSCT-7393

[11]. G. Bordel, M. Penagarikano, L. J. Rodríguez-Fuentes, A. Álvarez, and A. Varona, ‘‘Probabilistic kernels for improved text-to-speech alignment in long audio tracks,’’ IEEE Signal Process. Lett., vol. 23, no. 1, pp. 126–129, Jan. 2016.

[12]. A. Haubold and J. R. Kender, ‘‘Alignment of speech to highly imperfect text transcriptions,’’ in Proc. IEEE Multimedia Expo Int. Conf., Jul. 2007, pp. 224–227.

[13]. D. Yu and L. Deng, Automatic Speech Recognition: A Deep Learning Approach. London, U.K.: Springer, 2014.

[14]. T. J. Hazen, ‘‘Automatic alignment and error correction of human generated transcripts for long speech recordings,’’ in Proc. Interspeech, Sep. 2006, pp. 1–4.

[15]. G. E. Dahl, D. Yu, L. Deng, and A. Acero, ‘‘Context-dependent pretrained deep neural networks for large-vocabulary speech recognition,’’ IEEE Trans. Audio, Speech, Language Process., vol. 20, no. 1, pp. 30–42, Jan. 2012, doi: 10.1109/TASL.2011.2134090.

[16]. S. Hoffmann and B. Pfister, ‘‘Text-to-speech alignment of long recordings using universal phone models,’’ in Proc. Interspeech, Aug. 2013, pp. 1520–1524.

[17]. M. J. F. Gales, K. M. Knill, and A. Ragni, ‘‘Unicode-based graphemic systems for limited resource languages,’’ in Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP), Apr. 2015, pp. 5186–5190.

[18]. B. Safadi, M. Sahuguet, and B. Huet, ‘‘When textual and visual information join forces for multimedia retrieval,’’ in Proc. Int. Conf. Multimedia Retr., Apr. 2014, pp. 265–272

[19]. Kaldi a Toolkit for Speech Recognition. Accessed: Apr. 20, 2021. [Online]. Available: http://kaldi-asr.org/doc/

[20]. I. Gonzalez-Carrasco, L. Puente, B. Ruiz-Mezcua, and J. L. Lopez-Cuadrado, ‘‘Sub-sync: Automatic synchronization of subtitles in the broadcasting of true live programs in Spanish,’’ IEEE Access, vol. 7, pp. 60968–60983, 2019, doi: 10.1109/ACCESS.2019.2915581