

IMPORTANCE OF PROSODIC FEATURES IN LANGUAGE IDENTIFICATION

Niraj Kr. Singh
Research Scholar (M.Tech*)
Vivekananda Global University, Jaipur
Jaipur, India
Connect2vit@gmail.com

Anoop Singh Poonia
Professor, E & C Engineering
Vivekananda Global University, Jaipur
Jaipur, India
anoopsingh_25@rediffmail.com

I. INTRODUCTION

Abstract: *Earlier Researches in LID systems found that inclusion of prosodic features (alike, speech rate, Fundamental Frequency and Syllable timing) offered a little to develop the performance of their systems. Focused study on the utility of prosodic feature is attempted to evaluate parameters to behold the fundamental frequency and amplitude contours on syllable – by-syllable bases. The timing relationship is computed from the F0 whereas histogram of features (feature pairs) are collected through amplitude information followed by computations of loglikelihood ratio function for evaluating the unknown utterance (word/speech) in a pair wise discrimination task. The consequence reflects prosodic parameters to be useful in discriminating languages. Recent works in LID using prosody have proven appreciating results with better accuracy. This paper describes the importance of prosodic features in comparison with other possible features and methods for language identification.*

Keywords: *Prosody, Utterance, Modeling, Expectation Maximization, GMM.*

In the discussion of related work, we focus on previous work in sign language recognition. For coverage of gesture recognition, the survey in [24] is an excellent starting point. Other, more recent work is reviewed in [35]. Much previous work has focused on isolated sign language recognition with clear pauses after each sign, although the research focus is slowly shifting to continuous recognition. These pauses make it a much easier problem than continuous recognition without pauses between the individual signs, because explicit segmentation of a continuous input stream into the individual signs is very difficult. For this reason, and because of co articulation effects, work on isolated recognition often does not generalize easily to continuous recognition.

Erensthteyn and colleagues used neural networks to recognize finger spelling [6]. Waldron and Kim also used neural networks, but they attempted to recognize a small set of isolated signs [34] instead of finger spelling. They used Stokoe's transcription system [29] to separate the

handshape, orientation, and movement aspects of the signs. Kadous used Power Gloves to recognize a set of 95 isolated Auslan signs with 80% accuracy, with an emphasis on computationally inexpensive methods [13]. Grobel and Assam used HMMs to recognize isolated signs with 91.3% accuracy out of a 262-sign vocabulary. They extracted 2D features from video recordings of signers wearing colored gloves [9]. Braffort described ARGo, an architecture for recognizing French Sign Language. It attempted to integrate the normally disparate fields of sign language recognition and understanding [2]. Toward this goal, Gibet and colleagues also described a corpus of 3D gestural and sign language movement primitives [8]. This work focused on the syntactic and semantic aspects of sign languages, rather than phonology. Most work on continuous sign language recognition is based on HMMs, which offer the advantage of being able to segment a data stream into its constituent signs implicitly. It thus bypasses the difficult problem of segmentation entirely.

II. BACKGROUND

Speech recognition technology made major strides in the 1970s, thanks to interest and funding from the U.S. Department of Defense. The DoD's DARPA Speech Understanding Research (SUR) program, from 1971 to 1976, was one of the largest of its kind in the history of speech recognition, and among other things it was responsible for Carnegie Mellon's "Harpy" speech-understanding system. Harpy could understand 1011 words, approximately the vocabulary of an average three-year-old.

Over the next decade, i.e. 1980s, new approaches to understanding what people say, speech recognition vocabulary jumped from about a few hundred words to several thousand words, and had the potential to recognize an unlimited number of words. One major reason was a new statistical method known as the *HMM (Hidden Markov Model)*. As the time scope was limited and to be able to focus on more specific issues than HMM in general, the Hidden Markov Model toolkit (HTK) was used. HTK is a toolkit for building Hidden Markov Models (HMMs). HMMs can be used to model any time series and the core of HTK is similarly general-purpose.

III. PROSODY FOR LID

Speech is one of the oldest and most natural means of information exchange between human beings. We as humans speak and listen to each other in human-human interface. For centuries people have tried to develop machines that can understand and produce speech as humans do so naturally (Pinker, 1994 [20]; Deshmukh et al., 1999 [5]). Obviously such an interface would yield great benefits (Kandasamy, 1995,) [12]. Attempts have been made to develop vocally interactive computers to realise voice/speech recognition. In this case a computer can recognize text and give out a speech output (Kandasamy, 1995) [12]. Speech recognition can be defined as the process of converting an acoustic signal, captured by a microphone or a telephone, to a set of words (Zue et al., 1996 [36]; Mengjie, 2001 [17]). Automatic speech recognition (ASR) is one of the fastest developing fields in the

framework of speech science and engineering. As the new generation of computing technology, it comes as the next major innovation in man-machine interaction, after functionality of text-to-speech (TTS), supporting interactive voice response (IVR) systems.

IV. DISCUSSIONS

The implementation purposes the following methods were taken into practice : (a) Building the task grammar, (b) Constructing a dictionary for the models (c) Recording the data (d) Creating transcription files for training data (e) Encoding the data (feature processing) (f) (Re-) training the acoustic models (g) Evaluating the recognizers against the test data (h) Reporting recognition results. These are exhaustive representation of a LID system and do require more robustness for better performance.

REFERENCES

- [1] Gauvain, "Language identification using phonebased acoustic likelihoods," in Proc. ICASSP '94, vol. 1, Apr. 1994, pp. 293-296.
- [2] S. Kadambe and J. L. Hieronymus, "Language identification with phonological and lexical models," in Proc. ICASSP '95, vol. 5, May 1995, pp. 3507-3510.
- [3] R. J. D'Amore and C. P. Mah, "One-time complete indexing of text: Theory and practice," in Proc. Eighth Int. ACM Con\$ Res. Dev. Inform. Retrieval, 1985, pp. 155-164.
- [4] R. E. Kimbrell, "Searching for text? Send an N-gram!," Byte, vol. 13, no. 5, pp. 297-312, May 1988.
- [5] J. C. Schmitt, "Trigram-based method of language identification," US Patent 5 062 143, Oct. 1991.
- [6] M. Damashek, "Gauging similarity via N-grams: Language-independent text sorting, categorization, and retrieval of text," submitted for publication in Sei. T. A. Albina et al., "A system for clustering spoken documents," in Proc. Eurospeech '93, vol. 2, Sept. 1993, pp. 1371-1374.
- [7] Y. Yan and E. Barnard, "An approach to automatic language identification- '92, vol. 2, Oct. 1992, pp. 1007-1010. 1994, pp. 289-292.
- [8] S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," IEEE Trans Acoust, Speech, Signal Processing, vol. ASSP- 28, no. 4, pp. 357-366, Aug 1980.
- [9] D. B. Paul, "Speech recognition using hidden Markov models," Lincoln Lab. J., vol. < no. 1, p; 41-62, Spring 1990.
- [10] Y.K. Muthusamy, E. Barnard, and R.A. Cole, "Reviewing Automatic Language Identification", in IEEE Signal Processing Magazine, October 1994. Acquired: IEEE CD-ROM database. Usefulness: 10. Readability: 7.
- [11] Y.K. Muthusamy, R.A. Cole, B.T. Oshika, "The OGI Multi-Language Telephone Speech Corpus", in Proceedings International Conference on Spoken Language Processing 1992. <http://www.cse.ogi.edu/CSLU/corpora/mlts.html> Acquired: The WWW. Usefulness: 9. Readability: 10.
- [12] G.A. Constantinides, "A Framework for Evaluating Multilingual Systems", Surprise '96, http://www-dse.doc.ic.ac.uk/~nd/surprise_96/journal/vol1/gac1/article1.html. Acquired: My Brain. Usefulness: 10. Readability: 10. (of course!)
- [13] Y.K. Muthusamy, N. Jain, R.A. Cole, "Perceptual Benchmarks For Natural Language Identification", in Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing 1994.

- Acquired: IEEE CD-ROM database. Usefulness: 7. Readability: 7.
- [14] J.R. Deller, J.G. Proakis, J.H.L Hansen, "Discrete-Time Processing of Speech Signals", MacMillan, New York, 1993. Acquired: Friend. Usefulness: 7. Readability: 5. (Highly Mathematical)
- [15] W.A. Ainsworth, "Speech Recognition by Machine", Peter Peregrinus, London, 1988. Acquired: Main Library. Usefulness: 8. Readability: 9. (Nothing on multilingual stuff)
- [16] S.B. Davis and P. Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences", in IEEE Transactions on Acoustics, Speech and Signal Processing, Vol. ASSP-28, No.4, August 1980. Acquired: EE Library. Usefulness: 5. Readability: 5.
- [17] P. Mermelstein, "Automatic Segmentation of Speech", J. Acoust. Soc. Amer., Vol. 58, pp. 880-883, Oct 1975.
- [18] Y. K. Muthusamy, E. Barnard, and R. A. Cole, "Reviewing automatic language identification," IEEE Signal Processing Mag., vol. 11, no. 4, pp. 3341, Oct. 1994.
- [19] L. Riek, W. Mistretta, and D. Morgan, "Experiments in language identification," Lockheed Sanders, Inc., Nashua, NH, Tech. Rep. SPCOT- 91-002, Dec. 1991.
- [20] M. A. Zissman, "Automatic language identification using Gaussian mixture and hidden Markov models," in Proc. ICASSP '93, vol. 2, Apr. 1993, pp. 399402.
- [21] T. J. Hazen and V. W. Zue, "Automatic language identification using a segment-based approach," in Proc. Eurospeech '93, vol. 2, Sept. 1993, pp. 1303-1306.
- [22] M. A. Zissman and E. Singer, "Automatic language identification of telephone speech messages using phoneme recognition and n-gram modeling," in Proc. ZCASSP '94, vol. 1, Apr. 1994, pp. 305-308.
- [23] R. C. F. Tucker, M. J. Carey, and E. S. Paris, "Automatic language identification using sub-words models," in Proc. ZCASSP '94, vol. 1, Apr. 1994, pp. 301-304.
- [24] L. F. Lamel and J.-L. Gauvain, "Identifying non-linguistic speech features," in Proc. Euro speech '93, vol. 1, Sept. 1993, pp. 23-30.
- [25] Y. Muthusamy et al., "A comparison of approaches to automatic language identification using telephone speech," in Proc. Eurospeech '93, vol. 2, Sept. 1993, pp. 1307-1310.
- [26] Y. K. Muthusamy, R. A. Cole, and B. T. Oshika, "The OGI Multilanguage telephone speech corpus," in Proc. ZCASP '92, vol. 2, Oct. 1992, pp. 895-898.
- [27] D. Cimarusti and R. B. Ives, "Development of an automatic identification system of spoken languages: Phase I," in Proc. ICASSP '82, May J. T. Foil, "Language identification using noisy speech," in Proc. ICASSP '86, vol. 2, Apr. 1986, pp. 861-864.
- [28] F. J. Goodman, A. F. Martin, and R. E. Wohlford, "Improved automatic language identification in noisy speech," in Proc. ICASSP '89, vol. 1, 1982, pp. 1661-1663. . - May 1589, pp. 528-531.
- [29] R. B. Ives, "A minimal rule AI expert system for real-time classification of natural spoken languages," in Proc. Second Ann. Artificial Intel. Adv. Compute. Technol. Conf., Long Beach, CA, May 1986, pp. 337-340.
- [30] M. Sugiyama, "Automatic language recognition using acoustic features," in Proc. ICASSP '91, vol. 2, May 1991, pp. 813-816.
- [31] Y. K. Muthusamy and R. A. Cole, "Automatic segmentation and identification of ten languages using telephone speech," in Proc. ICSLP A. S. House and E. P. Neuburg, "Toward automatic identification of the language of an utterance. I. Preliminary methodological considerations," J. Acoust. Soc. Amer., vol. 62, no. 3, pp. 708-713, Sept. 1977. M. Savic, E. Acosta, and S. K. Gupta, "An automatic language identification system," in Proc. ICASSP '91, vol. 2, May 1991, pp. 817420.
- [32] S. Nakagawa, T. Seino, and Y. Ueda, "Spoken language identification by ergodic HMMs and its state sequences," Electron. Commun. Japan, Pt. 3, vol. 77, no. 6, pp. 70-79, Feb. 1994.