

# [Result of vlid using prlm approach english language essay](https://assignbuster.com/result-of-vlid-using-prlm-approach-english-language-essay/)

[](https://assignbuster.com/)[Linguistics](https://assignbuster.com/essay-subjects/linguistics/), [English](https://assignbuster.com/essay-subjects/linguistics/english/)

1M. E, COMMUNICATION SYSTEMS Student, 2Professor, Department of ECE, Dhanalakshmi Srinivasan Engineering College, PerambalurEmail-1muthupriyaece@gmail. com

## Abstract- Visual Language is identified from the movement of speech articulators of human using Automatic visual language identification and Automatic Lip reading techniques. By using this words spoken by the humans can be identified. It describes the datasets we recorded for the task, and describes the techniques and visual features weuse. Atpresentthistechniqueis implementedinSpeakerDependent Language Identification and it becomes ineffective due to the Noisy environment. We then extend the technique to Speaker Independent mode of operation.

## Indexterms-Activeappearance model (AAM), Automatic Lip reading, Visual LanguageIdentification(VLID), Visual Speech recognition.

## I INTRODUCTION

Speech can be Identified Visually from cues that are used by Humans in order to improve the Speech perception under noisy conditions [1], and using many visual features to recognize the various kinds of speech is only the main goal ofthe project [8]. If there is no any Audio signal during the process, then it is said to be lip reading. In the case of Deaf humans, thisLipreadingtechniques canbe implemented [4], So that, here the Lip movement is keenly observed [2] and then by applying Lip reading techniques the words can be identified as mentioned. LID is most commonly used technique for automaticallyidentifyingthelanguage from the speaker [6]. This Audio LID is now further used invisualmeans. Visuallanguage Identification is one of the advanced methodsofLanguageidentification technique (LID), this is used in many Applications such as Law-enforcement, On-linee-learning, Video-conferencing [11]. In this paper describes about the techniques that are implemented in the field of VLID. This paper is structure as follows: in section II, we give the relevant analysis of the language identificationmethods, including brief reviews of the primary Audio LID technique. Section III describes the techniques in the system; Section IV describes the Experiments and Discussion about the project. The Speaker independent language Identification [3] can be used in Automatic Lip reading technique [7] and it can be implemented for future work.

## II ANALYSIS

Languageidentificationisthe processofdeterminingwhichnatural language given content is in. Traditionally, identification of written language - aspracticed, for instance, in library sciencediscriminatory feature between languages as mentioned in [3],[6] and [7]. a) Phone- Based Tokenization: To exploit the difference in phonetic content between languagestoachievethelanguage discrimination. The contention here is that different languages have different rules regarding the syntax of phones, and this can be captured in a language model. Such techniques require the training of a phone recognizer, it usually comprises set of Hidden Markov models (HMMs), which are used to segment the input speech into a sequence of phones. has relied on manually identifying frequent wordsand lettersknowntobe characteristicofparticularlanguages. More recently, computational approachesSpeechMFCCfeatureextractionEnglishlanguage modelFrench language modelPr(En| EnEn, m)Pr(Fr| EnFn, m)have been applied to the problem, by viewinglanguageidentificationasa special caseof textcategorization, a Natural Language Processing approach that relies on statistical methods. Audio Language identification is a mature fieldofresearch, withmany successful techniques developed to achieve high levels of language discrimination with only a few seconds of test data. The main approaches make use of the phonetic and phonotactic characteristics of languages whichareproventobe identifiable

## Fig1. Diagram for phone recognition

## followed by Language modeling approach to audio LID

Hereinthe(fig. 1), asinglephone recognition system is used to tokenize an utterance using a shared phone set, trained using one language. Phone recognition followed by language modeling approach is used; Phonotactics is the feature of language used for discrimination. Here different languages have different rules regarding the syntax of phones, and this can be captured in a language model. b) Gaussian Mixture Model Tokenization: The tokenization subsystem within the LID system is usually applied at a phone level. Gaussian mixture model (GMM) is trained for each language [9]. Each GMM can be considered to be an acoustic dictionary of sounds, witheachmixturecomponent modeling a distinct sound from the training data. c) From language-specific acoustic data. Each GMM can be considered to be an acoustic dictionary of sounds, with each mixture component modeling a distinct sound from training data. Here the component becomes the token of frame. For a stream of input frames, a stream of componentindices willbe produced, on which language modeling followed by back-end classification can be performed in audio LID as mentioned in [7]&[8].

## III. TECHNIQUES

a) LIP READINGLip reading, also known as lip reading or speech reading, is a technique of understandingspeechbyvisually interpreting the movements of the lips, face and tongue with information provided by the context, language, and any residual hearing. This is because each speech sound (phoneme) has a particular facial andmouth position (viseme), although many phonemes share the same viseme and thus are impossible to distinguish from visual information alone [13]. When a normal person speaks, the tongue moves in at least three places (tip, middle and back), and the soft palate rises and falls. Consequently, sounds whose place of articulation is deep insidethemouthorthroatarenot detectable, such as glottal consonants. Voiced and unvoiced pairs look identical, (inAmericanEnglish); likewise for nasalization.

## Fig2. Lip reading

b) Active Appearance ModelHere for speech recognition we use AAM (for appearance) and ASM (for shape). In independent experiments we use AAM features. However, in our earlier, speaker dependent experiments, we used ASM features; they also provide good language discrimination. To construct an AAM, a selection of training images is marked with a number of points that identify the features of interest on the face. AAM appearance is computed for each training image is shape normalized by warping it from the labeled feature points, to the mean shape. Our implementation of the AAM uses the RGB color space. The pixel intensities within the mean shape are concatenated, and the vectors representing each color channel are then concatenated. We use the inverse compositional project- out algorithm [10] to track landmark positions over a sequence of video frames. Thisalgorithmiterativelyadjuststhe landmarkpositionsonanimageby minimizing the error between the mean appearance and the appearance contained by the current landmarks, warped to the mean shape.

## Fig3. Mean and first three modes of variation of the appearance component of an AAM

## IVEXPERIMENTSAND DISCUSSION

Herethistechniquecanbe implemented by observing the movement of speech articulators such as lips jaws and teeth andfurtherthelanguagebeing spoken and the text is identified. Languages such as English, French and German are focused for the language identification in this paper. In this moduledemonstrated that VLID is possible in both speaker-dependent and independent cases, and that there is sufficient information presentedon thelips to discriminate between two or three languages using these techniques, despitethelowphone recognition accuracies that were observed. AAM features are well separated between speakers, meaning that there is no correspondence between the feature vectors for each speaker. AAM features are well separated between speakers, meaning that there is no correspondence between the feature vectors for each speaker. Based upon the above results and experiments the language has been identified inthismodelbaseduponthedataset construction.

## Fig4. Result of VLID using PRLM approach

## Fig5. Result of VLID using ASM features

## Fig6. Result of VLID using AAM features

## V CONCLUSION

In this paper, we have presented an account of initial research into the task of VLID. We have developed two methods forlanguageidentificationofvisual speech, based upon audio LID techniquesthat use language phonology as a feature of discrimination: an unsupervised approach that tokenizes ASM feature vectors using VQ, and a supervised method of visual triphone modelling using AAM features. Wehavedemonstrated thatVLIDis possible in both speaker-dependent and independentcases, andthat thereis sufficient information presented on the lips to discriminate between two or three languages using these techniques, despite the low phone recognition accuracies that we observed. Throughout, we have taken pains to ensure that the discrimination between languages we have obtained is genuine and not based on differences in the recording or the speakers. Apart from one three-language discrimination task described in Section IV, thisresearchhasfocusedon discriminating between two languages. In thefuture, thenumberoflanguages included in the system should be increased todeterminehow well this approach generalizes when the chance of language confusion is higher. Groups of phonetically similar languages could be added to see if they are more confusable than those with differing phonetic characteristics, as well as tonal languages. Phonotactics are not the only aspect oflanguagewhichcanbeusedto differentiate between them. Further workintoVLIDcouldthereforefocuson incorporating both of these additional languagecuesandevaluatingtheir contribution to language discrimination.

## VI REFERENCES

[1]Q. Summerfield, " Lipreading and audio-visualspeechperception," Philosophy. Trans.: Biol. Sci., vol. 335, no. 1273, pp. 71–78, 1992.[2]G. Potamianos, C. Neti, G. Iyengar, and E. Helmuth, " Large-vocabulary audio-visual speech recognition by machines and humans," in Proc. Euro speech ’01, 2001, pp. 1027–1030.[3]L. Liang, X. Liu, Y. Zhao, X. Pi, andA. Nefian," Speaker independentaudio-visual continuous speech recognition," in Proc. IEEE Int. Conf. Multimedia Expo (ICME), 2002, vol. 2, pp. 25–28.[4]I. AlmajaiandB. Milner, " Enhancingaudiospeechusing visual speech features," in Proc. Interspeech ’09, 2009, pp. 1959–1962.[5]C. Bregler and Y. Konig, " Eigen lips" for robustspeech recognition," inProc. IEEEInt. Conf. Acoust., Speech, SignalProcess. Apr. 1994, vol. 2, pp. 669–672.[6]I. Matthews, T. Cootes, J. Bangham, S. Cox, and R. Harvey, " Extraction of visual features for lipreading," IEEE Trans. Pattern Anal. Mach. Intell. (PAMI), vol. 24, no. 2, pp. 198–213, Feb. 2002.[7]M. Zissman, " Comparison of four approaches to automatic language identification of telephone speech," IEEETrans. SpeechAudio Process., vol. 4, no. 1, pp. 31–44, Jan. 1996.[8]Y. Muthusamy, E. Barnard, and R. Cole," Reviewingautomatic languageidentification," IEEE Signal Process. Mag., vol. 11, no. 4, pp. 33–41, Oct. 1994.[9]P. A. Torres-Carrasquillo, E. Singer, M. A. Kohler, R. J. Greene, D. A. Reynolds, " Approaches to language identificationusing Gaussianmixturemodelsand shifted delta cepstral features".[10]Mathewsands. Baker" Activeappearance models revisited" in2004 vol pp. 135-164.