DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

final paper on

Applications of Decision Trees in Phonetic Recognition

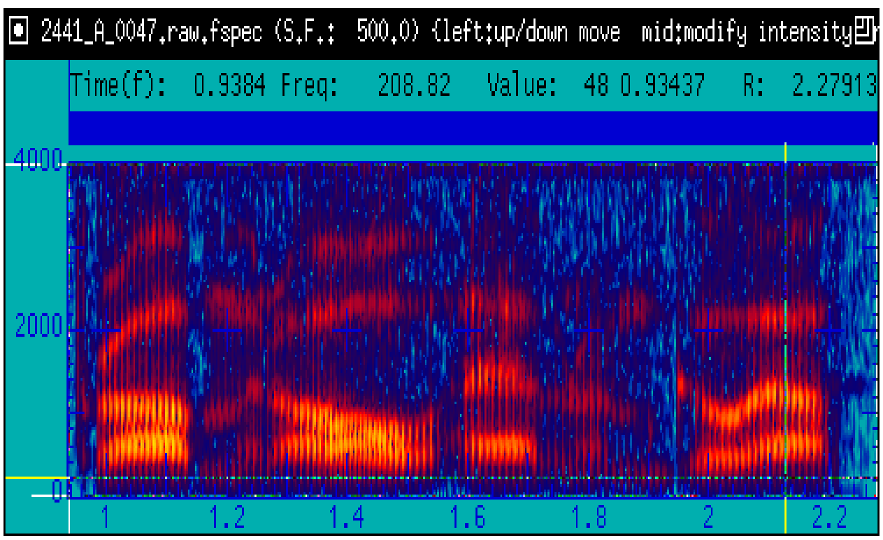
submitted to fulfill the requirements for

ECE 8993: Fundamentals of Speech Recognition

submitted to:

Dr. Joseph Picone
Department of Electrical and Computer Engineering
413 Simrall, Hardy Rd.
Mississippi State University
Mississipppi State, Mississippi 39762

May 9, 1998



submitted by:

Julie Ngan
MS in Computer Engineering Candidate
Department of Electrical and Computer Engineering
Mississippi State University
429 Simrall, Hardy Rd.
Mississippi State, Mississippi 39762
Tel: 601-325-8335, Fax: 601-325-3149
Email: ngan@isip.msstate.edu

ABSTRACT

Recent progress on conversational speech recognition has been painstakingly slow, and has resulted in measurable reductions in word error rate primarily through intricate pronunciation modeling and hand-tailoring of the lexicon.

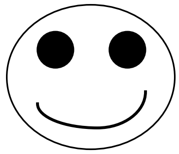


Figure 1. The meaing of life is depicted here, and explained in a full sentence.

1. INTRODUCTION

Recent progress on conversational speech recognition has been painstakingly slow, and has resulted in measurable reductions in word error rate primarily through intricate pronunciation modeling and hand-tailoring of the lexicon. The primary problem seems to stem from poor acoustic-level matching as a result of a high degree of variability in pronunciations. Three common approaches to overcome this problem are (1) prediction of common alternate pronunciations and incorporation of these as additional paths in the acoustic models, (2) use of acoustic units that specifically model multi-word phrases, and (3) reestimation of the acoustic models in such a way that the models automatically learn such alternate pronunciations.

A really cool equation, that explains everything in life, is as follows:

$$y = mx + b \tag{1}$$

A second equation, that is equally profound, is:

$$y = ax^2 + bx + c \tag{2}$$

I think we need a figure here to explain why this equation is the bomb. See Figure 1, which explains the meaning of life in a block diagram.

The first approach suffers from related problems of intelligent integration of language model and acoustic model scores. The last approach, which is featured in this white paper, requires robust training algorithms that can converge to meaningful solutions in the presence

1.1. My First Sub-Heading

I think it is about time to introduce a simple table. In Table X, we see how a table should look. Recent progress on conversational speech recognition has been painstakingly slow, and has resulted in measurable reductions in word error rate primarily through intricate pronunciation modeling and hand-tailoring of the lexicon.

1.2. My Second Sub-heading

Recent progress on conversational speech recognition has been painstakingly slow, and has resulted in measurable reductions in word error rate primarily through intricate pronunciation modeling and hand-tailoring of the lexicon.

1.3. My Third Sub-Heading

Recent progress on conversational speech recognition has been painstakingly slow, and has resulted in measurable reductions in word error rate primarily through intricate pronunciation modeling and hand-tailoring of

Date	Time	Value
3	4	7
6	6	6
5	5	5

Table 1. This table is really important, so we spend a lot of time explaining it.

the lexicon.

1.4.1. A Really Deep Theory

Recent progress on conversational speech recognition has been painstakingly slow, and has resulted in measurable reductions in word error rate primarily through intricate pronunciation modeling and hand-tailoring of the lexicon.

2. DECISION TREES

Recent progress on conversational speech recognition has been painstakingly slow, and has resulted in measurable reductions in word error rate primarily through intricate pronunciation modeling and hand-tailoring of the lexicon.

3. APPLICATIONS IN LVCSR

Recent progress on conversational speech recognition has been painstakingly slow, and has resulted in measurable reductions in word error rate primarily through intricate pronunciation modeling and hand-tailoring of the lexicon.

4. EXPERIMENTAL SUPPORT

Recent progress on conversational speech recognition has been painstakingly slow, and has resulted in measurable reductions in word error rate primarily through intricate pronunciation modeling and hand-tailoring of the lexicon.

5. SUMMARY

Recent progress on conversational speech recognition has been painstakingly slow, and has resulted in measurable reductions in word error rate primarily through intricate pronunciation modeling and hand-tailoring of the lexicon.

6. ACKNOWLEDGEMENTS

Recent progress on conversational speech recognition has been painstakingly slow, and has resulted in measurable reductions in word error rate primarily through intricate pronunciation modeling and hand-tailoring of the lexicon.

REFERENCES

[1] A. Ganapathiraju, *et. al.*, “Syllable — A Promising Recognition Unit for LVCSR,” *Proceedings of the 1997 IEEE Automatic Speech Recognition and Understanding Workshop*, pp. 207-214, Santa Barbara, California, USA, December 1997.

[2] G. Doddington, et. al., “Syllable-Based Speech Recognition.” *WS’97 Technical Report*, Center for Language and Speech Processing, Johns Hopkins University, Baltimore, Maryland, USA, December 1997.

[3] J. Hamaker, A. Ganapathiraju, J. Picone, and J. Godfrey, “Advances in Alphadigit Recognition Using Syllables,” to be presented at the *IEEE International Conference on Acoustics, Speech, and Signal Processing*, Seattle, Washington, USA, May 1998.

[4] J. Whittaker, *Graphical Models in Applied Multivariate Statistics*, John-Wiley & Sons, New York, New York, USA, 1990.

[5] D. Heckerman et.al, “Probabilistic Independence Networks for Hidden Markov Probability Models,” *Technical Report MSR-TR-96-03*, Microsoft Research, Redmond, Washington, USA, June 1996.

[6] S. Chen et. al., “IBM’s LVCSR System for Transcription of Broadcast News used in the 1997 HUB4 English

Evaluation," To appear in the *Proceedings of the DARPA Broadcast News Transcription and Understanding Workshop,* Lansdowne, Virginia, USA, Feb. 8-11 1998.

[7] A. Stolke, S. Omohundro, "Best-first Model Merging for Hidden Markov Model Induction," *Technical Report TR-94-003,* ICSI, University of California, Berkley, California, USA, March 1994.

[8] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press, San Diego, California, USA, 1990.

[9] N. Kumar, "Investigation of Silicon-Auditory Models and Generalization of LDA for Improved Speech Recognition," *Ph.D. Thesis*, Johns Hopkins University, Baltimore, Maryland, USA, May 1997.

[10] C. Cortes, V. Vapnik, "Support Vector Networks," *Machine Learning*, vol. 20, pp. 273-297, September 1995.

[11] V. Vapnik, *Estimation of Dependencies Based on Empirical Data*, Springer-Verlag, New York, New York, USA, 1982.

[12] V. Vapnik, *The Nature of Statistical Learning Theory,* Springer-Verlag, New York, New York, USA, 1995.

[13] S. Gunn, "Support Vector Machines," *http://www.isis.ecs.soton.ac.uk/research/svm*, Image, Speech and Intelligent Systems Research Group, University of Southhampton, March 1998.

[14] P. Knirsch, "Demonstration 2-D Pattern Recognition Applet," *http://svm.research.bell-labs.com/SVT/SVMsvt.html*, Advanced Information Systems Engineering, Lucent Technologies, March 1998.

APPENDIX A

Recent progress on conversational speech recognition has been painstakingly slow, and has resulted in measurable reductions in word error rate primarily through intricate pronunciation modeling and hand-tailoring of the lexicon.