

Practical Speech Enhancement in Babble using Non-negative Matrix Factorisation

Ashley Gillman

James Cook University

May 22, 2014

Partner: Bryan Quill

Supervisor: Dr. Owen Kenny

What is Babble?

And why is it so interesting?

Type of Noise	No. of Speakers	Definition	Steady-state / Time-varying
Competing Speaker	2	There are two speakers present.	Time-varying
Babble	3 - 6	Individual speakers can be heard and at times, individual words can also be heard.	Time-varying
Speech-shaped (Crowd)	7+	A diffused background rumble, where individual conversations or speakers are not distinguishable.	Steady-state

A bit of History...

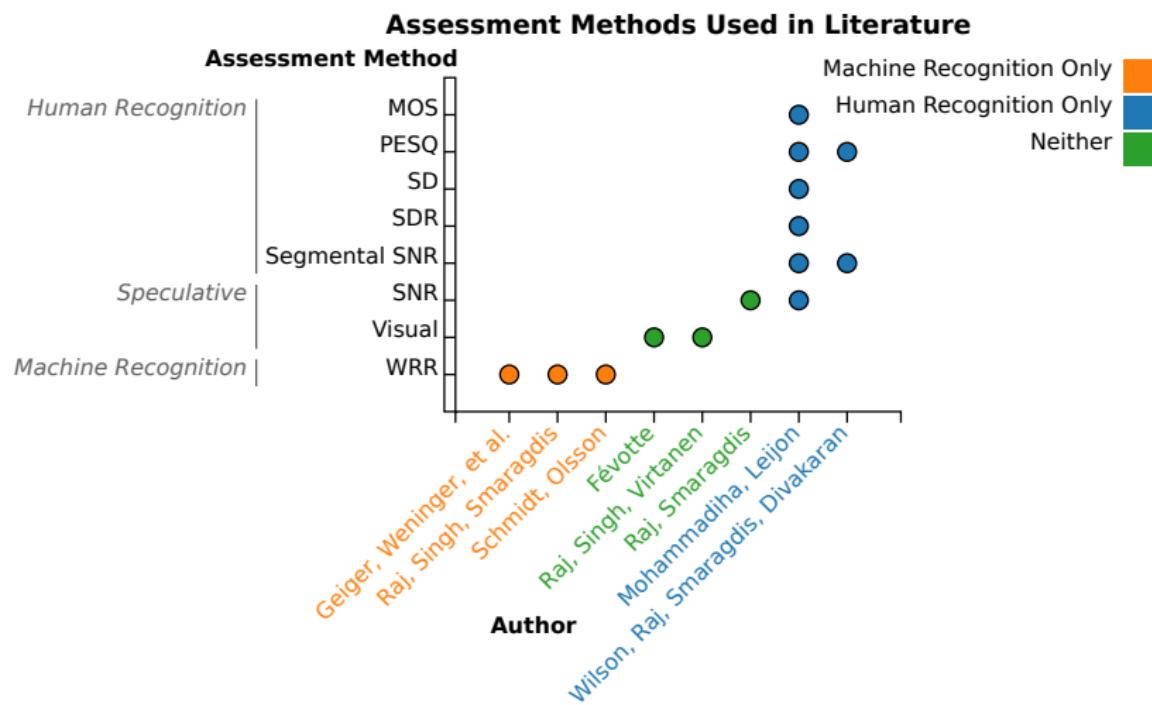
The cocktail party effect



By David Shankbone (Own work) [CC-BY-3.0 (<http://creativecommons.org/licenses/by/3.0>)], via Wikimedia Commons

Introduction

The first issue:



Research Question 1

“Are good enhancement algorithms effective for both human listeners and machine listeners?”

Can a generic and practical speech enhancement algorithm find application in signal enhancement and ASR?

Evaluation Measures

- ▶ Mean Opinion Score (MOS) - *Human listener*
- ▶ Perceptual Evaluation of Speech Quality (PESQ) - *Human listener*
- ▶ Automated Speech Recognition Word Recognition Rate (ASR-WRR) - *Machine listener*

Evaluation Measures

- ▶ Mean Opinion Score (MOS) - *Human listener*
- ▶ Perceptual Evaluation of Speech Quality (PESQ) - *Human listener*
- ▶ Automated Speech Recognition Word Recognition Rate (ASR-WRR) - *Machine listener*

Evaluation Measures

- ▶ Mean Opinion Score (MOS) - *Human listener*
- ▶ Perceptual Evaluation of Speech Quality (PESQ) - *Human listener*
- ▶ Automated Speech Recognition Word Recognition Rate (ASR-WRR) - *Machine listener*

Introduction

○○

Research Question 1

○
○
○

Research Question 2

●○
○
○○○

Research Question 3

○
○
○

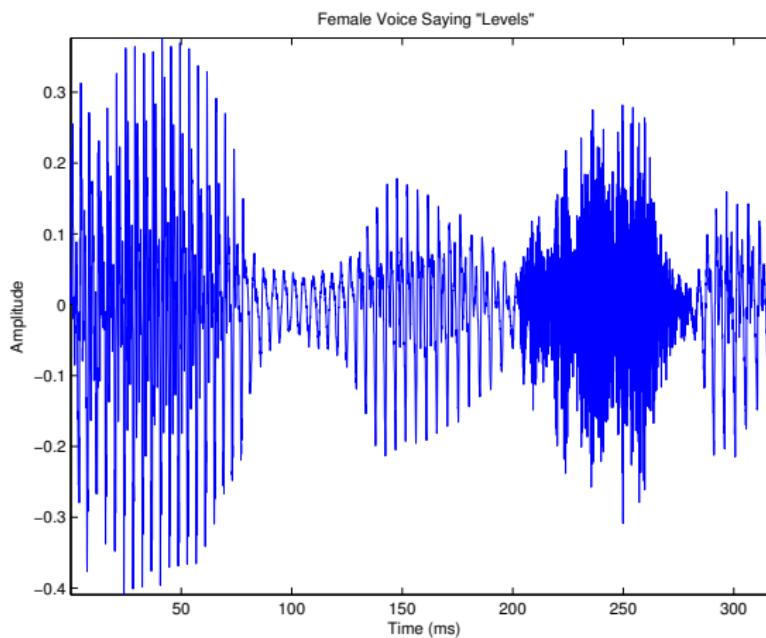
Progress

Conclusion

○○○

Introduction

Crash Course in Spectrograms



Introduction

○○

Research Question 1

○○○

Research Question 2

●○○○

Research Question 3

○○○

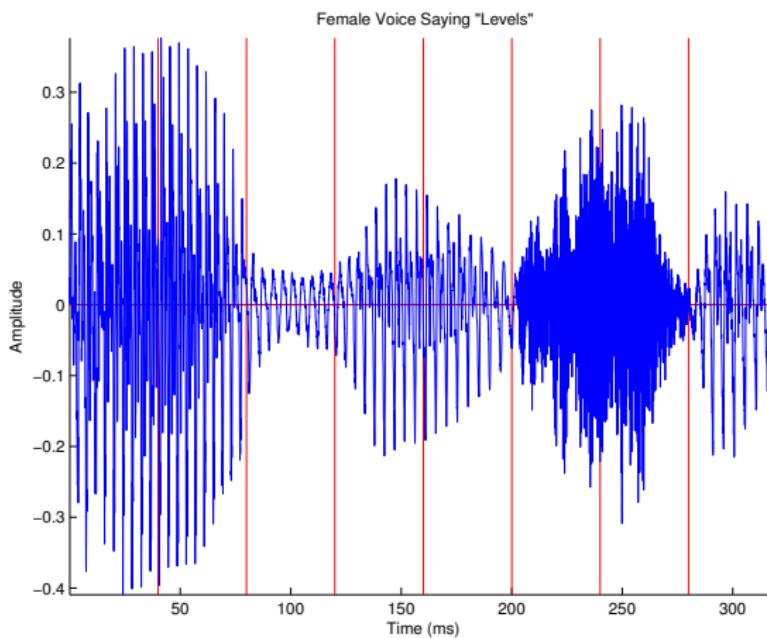
Progress

Conclusion

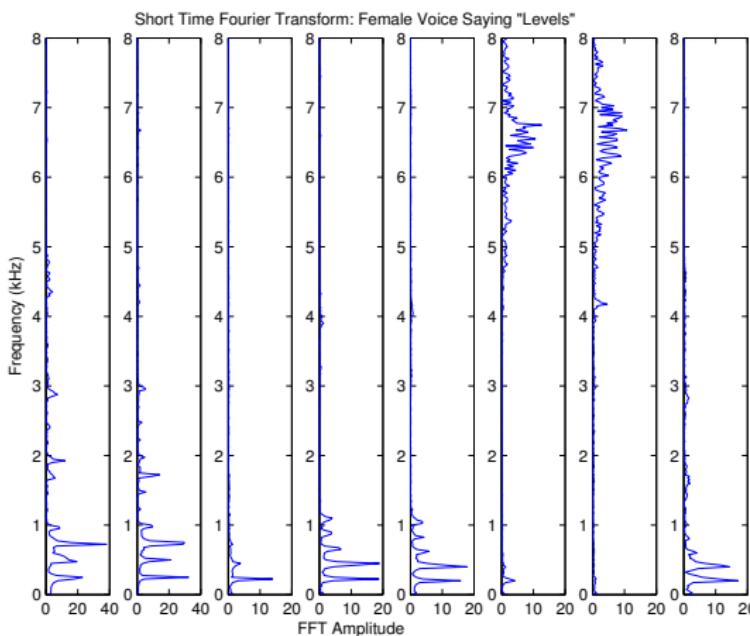
○○○

Introduction

Crash Course in Spectrograms

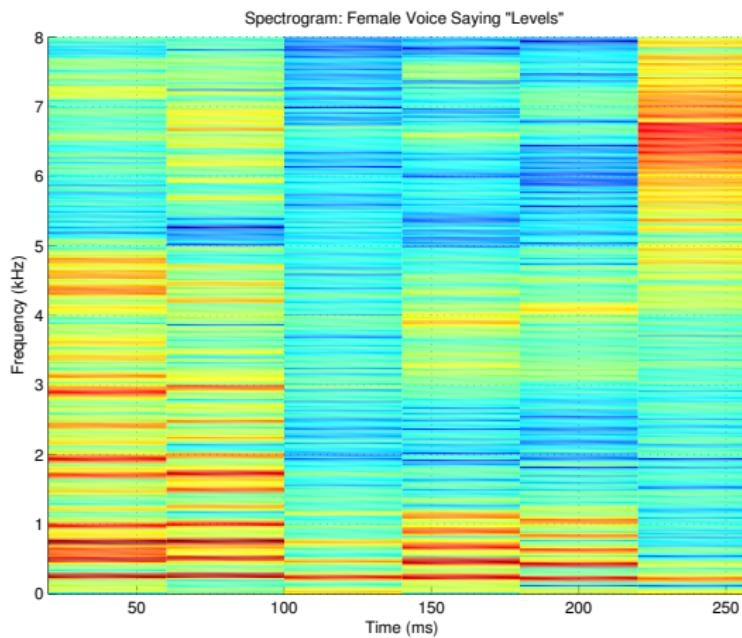


Crash Course in Spectrograms



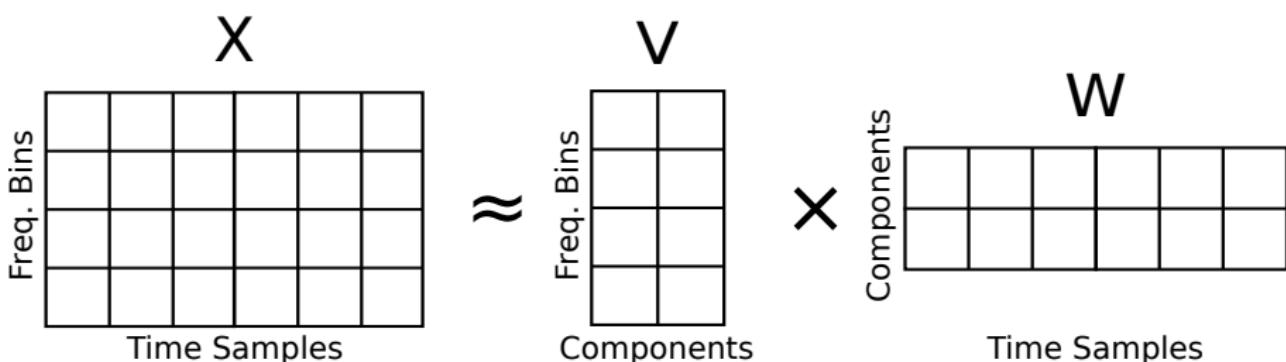
Introduction

Crash Course in Spectrograms



Introduction

Non-negative Matrix Factorisation



Adapted from Qwertysus (Own work) [CC-BY-SA-3.0 (<http://creativecommons.org/licenses/by-sa/3.0>)], via Wikimedia Commons

Definition

Research Question 2

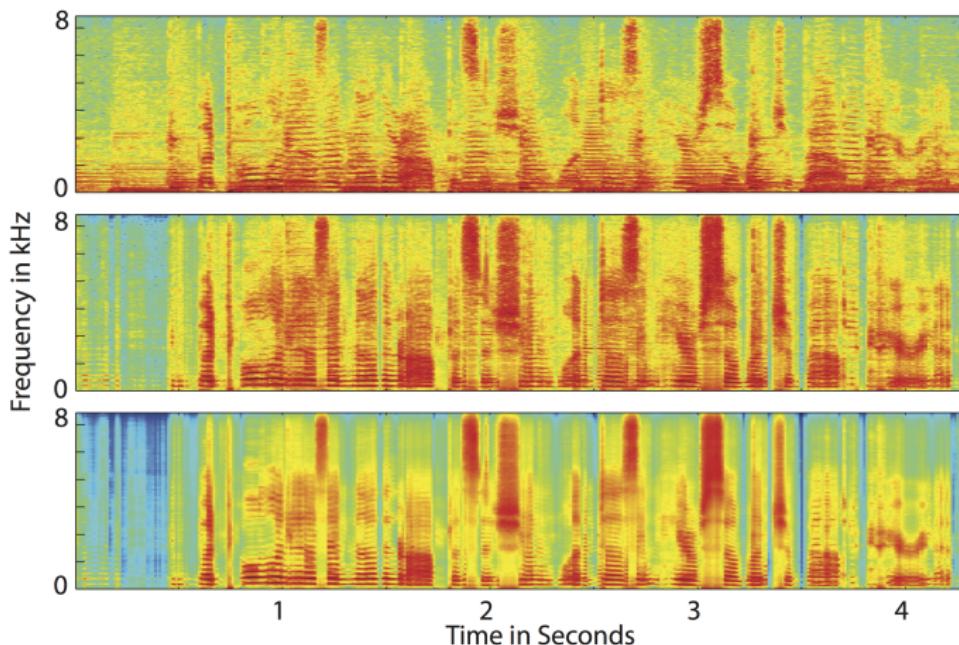
“Can the results be improved by modifying algorithms to concentrate the focus on recognition of the desired speakers voice?”

Phonemes

Phrase	“Examine a sentence.”
Words	“Examine” + “a” + “sentence”
Syllables	/Ex.am.ine/ + /a/ + /sen.tence/
Phonemes	/ɪ g z æ m ɪ n/ + /ə/ + /s ε n t ə n s/

Methodology

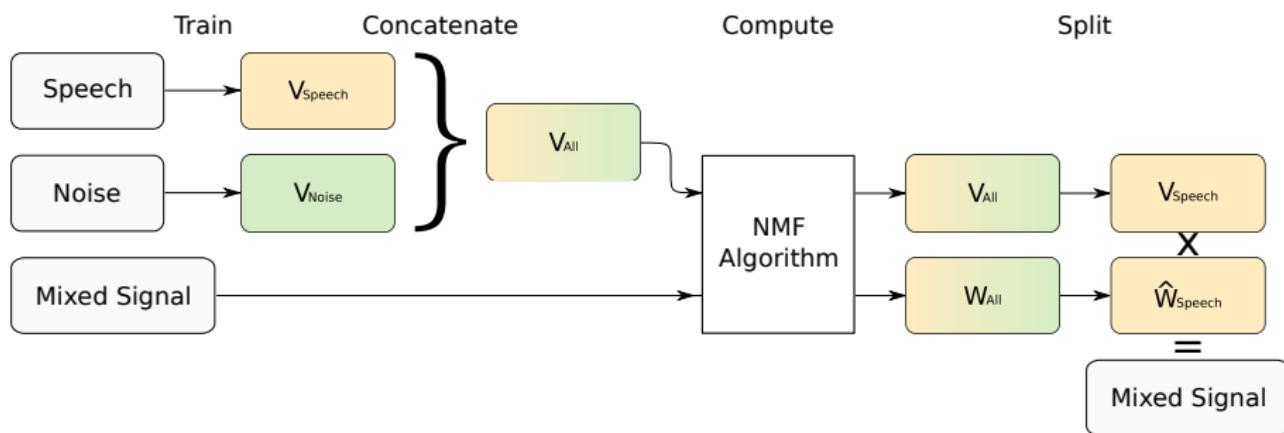
Previous Work



B. Raj, R. Singh, and T. Virtanen, "Phoneme-Dependent NMF for Speech Enhancement in Monaural Mixtures," in INTERSPEECH, 2011, pp. 1217-1220.

Methodology

Algorithm



Definition

Research Question 3

“Can the practicality of existing algorithms be improved to allow applications in end-user hardware?”

Introduction

○○

Research Question 1

○
○
○

Research Question 2

○○
○
○○○

Research Question 3

○
●
○

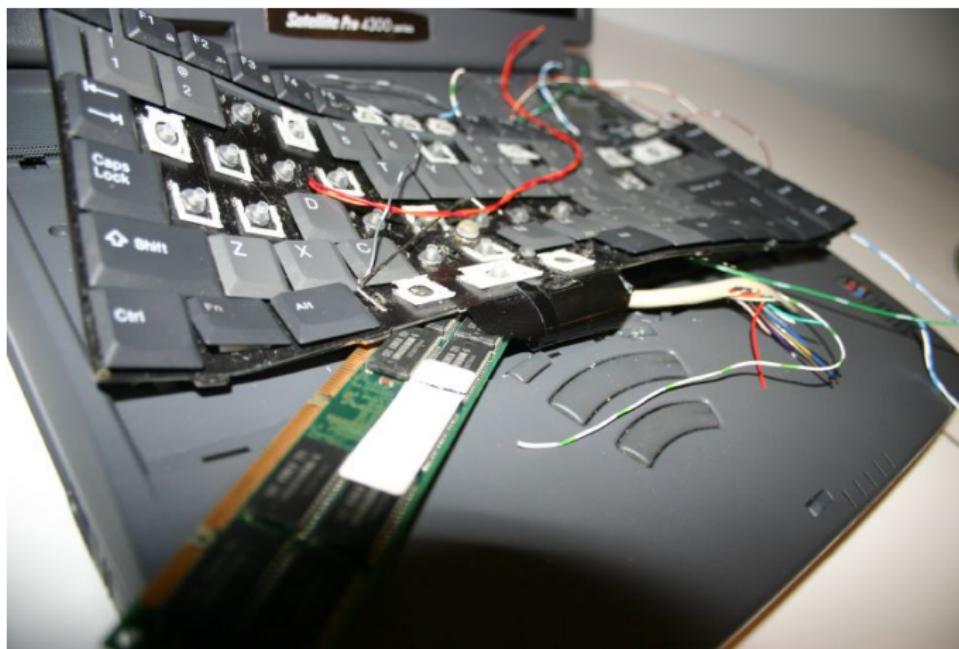
Progress

Conclusion

○○○

Need

Why is this necessary?



"Present" by S B [CC-BY-2.0 (<http://creativecommons.org/licenses/by/2.0>)], via flickr

Simplified voice training

Training requirements:

$$10 \text{ samples} \div (5 \frac{\text{phonemes}}{\text{word}} \times 140 \frac{\text{words}}{\text{minute}} \div 40 \text{ phonemes} \times 20\% \text{ usability rate}) = 2.9 \text{ mins}$$



"Reading e-mail" by Toms Bauģis [CC-BY-2.0 (<http://creativecommons.org/licenses/by/2.0/>)], via flickr

*Flopsy,
Mopsy,
and Cotton-tail*

*were very jealous of Peter's new speech
recognising computer*

Milestones and Progress

- ✓ Identify Problem
- ✓ Literature Review
- ✓ Develop Solution
- ✗ Implement Solution
- ✗ Test Solution

Plan of attack

- ▶ Verify these algorithms will be applicable to a broad range of practical situations, both for human and machine listeners.
- ▶ Verify phoneme-dependent NMF algorithms can improve speech in babble.
- ▶ Verify phoneme-dependent NMF algorithms can reduce training requirements, thereby rendering a more practical system.

Plan of attack

- ▶ Verify these algorithms will be applicable to a broad range of practical situations, both for human and machine listeners.
- ▶ Verify phoneme-dependent NMF algorithms can improve speech in babble.
- ▶ Verify phoneme-dependent NMF algorithms can reduce training requirements, thereby rendering a more practical system.

Plan of attack

- ▶ Verify these algorithms will be applicable to a broad range of practical situations, both for human and machine listeners.
- ▶ Verify phoneme-dependent NMF algorithms can improve speech in babble.
- ▶ Verify phoneme-dependent NMF algorithms can reduce training requirements, thereby rendering a more practical system.

Introduction

○○

Research Question 1

○
○
○

Research Question 2

○○
○
○○○

Research Question 3

○
○
○

Progress

Conclusion

○●○

No more “cocktail party problem”



"I can't hear you" by April Pink [CC-BY-SA-2.0 (<http://creativecommons.org/licenses/by-sa/2.0>)], via flickr

Introduction

oo

Research Question 1

o
o
o

Research Question 2

oo
o
ooo

Research Question 3

o
o
o

Progress

oo

Conclusion

oo●

“Questions?”

