Learning to Recognize Talkers from Natural, Sinewave, and Reversed Speech Samples

Presented by:
Pankaj Rajan
Graduate Student, Department of Computer Sciences.
Texas A&M University, College Station
Agenda

• Introduction
• Background
• Experiments
• Discussion
What the paper is about?

• How a listener recognizes the unfamiliar talker?

• Different talker specific properties of the speech signal to perceptual learning
Traditional View: Word and Talker Recognition

- Features for Speech perception and Talker recognition are different.

- As per Traditional View:
  - Vocal quality of Individual is represented by Features that are linguistically irrelevant.
  - For perception talker specific attributes are separated from phonetic representation.
Alternative to **Traditional view**

- Experiments have revealed relation between Linguistic and Indexical perception.

  - Evidence!!! *Experiments show that*
    
    - Listeners recognize phonemes and words faster and accurate when spoken by same speaker.
    - Talker attributes cannot be ignored.
    - Linguistic perception preserve talker specific attributes which become part of long term memory.

Experiments by Nygaard In *which he found that the listeners familiar with the talker were able to transcribe the sentences more easily*
Conclusions of Alternative view

- Talker properties and the linguistic form are associated with same single memory system.

- Linguistic and Indexical attributes are transmitted in parallel.
**Remez Experiment**

**AIM:** To test hypothesis that phonetic representations represent both talker characteristics and words.

- Listener familiar with the talker were presented the Sinewave representation of the natural speech for individual recognition.
- Listeners were able to recognize the talker and hence demonstrated that perception of talker characteristics is present in talker specific phoneme utterances.

**Sine Waves:** Acoustic structure of the original utterance that lack fine grained acoustic details of natural speech but evoke:

1. impression of segmental phonetic attributes and
2. Vocal quality like Fundamental frequency, broadband resonances, harmonic structure.
Spectrogram Representation
Present Work

- To establish if learners can perceive speaker in absence of acoustic vocal quality parameters.
- To know about the feature structure of the talker that develops during perceptual training.
- To know the extent to which the attributes of the talker are transferred to the type of signal.
5 Experiments:

**Experiment 1 and 2**
- Aimed to investigate if listener would recognize a talker from samples featuring phonetic segmental properties at the cost of qualitative aspects.

**Experiment 3, 4, 5**
- Aimed to explore how listeners would categorize unfamiliar talker when presented with qualitative aspects at the cost of segmental phonetic and lexical properties
Experiment
1 and 2
Experiments:

• Both Sinewave and the natural speech samples were used to train the listeners.
• Sentences were recorded in an audiotape in a soundproof booth and were low passed filtered at 4.5 kHz and sampled at 10 kHz.
• Subjects were trained to extent of 70% accuracy to identify the talker and training was conducted for several days.
• Subjects were encouraged to pay attention to the talker attributes rather than content of message.
Continued.....

- There was a familiarization phase, to reinstate correspondence between sine wave tokens and talkers name.
- It was followed by the Generalization test.
Experiment 1: Learning to identify talkers with Sinewave Sentences

Training:
- Listeners were trained in a quite room to identify 10 speakers of Sinewave utterances.

Analysis:
- Listeners were able to identify speakers and accuracy increases day by day.
- One way ANOVA revealed significant effect of talker identity on recognition performance.

Generalization Performance/Test:
- Half of the subjects were first presented with natural speeches to be identified before Sinewave and other half was presented in reversed order.

Results:
- Talker specific knowledge acquired during Sinewave training generalized to novel natural and Sinewave sentences.
Listener can identify talker with the phonetic attributes present in the Sinewave in absence of traditional qualitative attributes of vocal sound production.

It is perceptual discriminability of talker in the set that is source of differences in identification.

**Interesting Observation**

Individual listener differed in their ability to identify the talkers. However, reason was not known.

**Conclusion:**

Individual attributes are carried by the segmental phonetic properties in addition to vocal timber.

Linguistic and individual attributes can be represented by common representation code.
Figure 1. Vocal fold identification performance for Subjects 1-8 (S1-S8) in a function of training days and vocal use in Experiment 1.
Last day of Training Experiment 1
Generalized test

Natural Speech Generalization Performance

Sinewave Replica Generalization Performance
Experiment 2: Learning to identify talkers from natural speech

**Aim:**
- To find if similar generalization as observed in previous experiment can be obtained using natural speech training

**Training:**
- Subjects were trained on the natural sentences
- Similar to training in the Experiment 1.

**Analysis**
- Listeners were able to learn from natural speech very fast.
- It indicates that natural speech provides listeners with salient sample of each talker’s indexical attributes.
Generalized performance:
  - Similar to the first Experiment.

Results:
  - Listener’s ability to understand the speech was 88% for natural speech generalization test but 27% for Sinewave generalization test.

Conclusions
  - Indexical knowledge acquired during the training with the natural speech doesn't generalized to Sinewave utterances.
Last day Training Results - Exp: 2

Natural Speech Training Performance

Percent Correct

TALKER

F1 F2 F3 F4 F5 M1 M2 M3 M4 M5
Generalization Test
Experiment

3, 4, 5
**Experiment 3  Recognizing an unintelligible talker**

**AIM:**
- To test hypothesis that subjects use glottal source quality rather than fine grained phonetic properties during natural speech training.

**Training:**
- Listeners were trained using natural speech.

**Generalization Performance:**
- Generalization test samples were reversed natural speech and sinewave samples.

**Analysis:**
- Subjects were able to identify talker from reversed speech samples.
- Performance on the Sinewave samples was far poorer.
Continued.....

**Results**
- Qualitative attributes are prominent perceptual attributes during training with natural speech samples.
- Though results were poorer on Sinewave samples, correlation analysis shows that listeners do encoded some fine grained phonetic attributes for the recognition.

**Conclusion**
- Under Normal circumstances listeners naturally encode talkers using mixture of different properties, relying more on qualitative characteristics while also using the phonetic details
Generalization test: Exp:3

![Generalization Performance Chart](chart.png)
Experiment 4  Getting to know unintelligible talker

Aim:
- To test the sufficiency of qualitative aspects of speech in relative absence of many phonetic and lexical impressions.
- Ability to learn from reversed samples is tested.

Training
- Training material were reversed speech samples.

Generalization Process
- The generalized material are reversed speech samples and sine wave speeches sentences.

Analysis
- Listener learned to identify the individuals from reversed speech samples.
- Rate was intermediate between training with natural speech and with sine wave replica.
Continued.....

**Results**

- Reveres speech training was highly correlated with reverse speech talker identification.
- No correlation between training and Sinewave talker identification.

**Conclusion**

- The features with which listeners become familiar during training did not correspond to indexical attributes available in Sinewave replica.
Last day Training Performance

Reversed Speech Training Performance

Percent Correct

TALKER

F1  F2  F3  F4  F5  M1  M2  M3  M4  M5
Generalization test Exp: 4
Experiments Robustness of Qualitative Indexical attributes

Aim
- To test that the qualitative attributes available in reversed speech samples match those of natural speech.

Training
- Reversed natural speech samples.

Generalized performance
- Generalized test material were natural speech samples and Sinewave replicas.

Analysis
- Results were identical with the previous experiment.

Results
- Similar to the previous experiment.
- Weak Correlation between reversed speech training and Sinewave generalization.
Generalization test Exp: 5
Discussion:

- The listeners exploit various cues to perceive the talker.
- In some circumstances global qualitative attributes are used to perceive talker while in some circumstances, phonetic attributes alone can be used.