Effects of Perceived Gender on Prominence Perception

Tim Mahrt

Workshop de lancement du projet SIRL

Laboratoire Parole et Langage

20 - 11 - 2015

Overview

- Introduction
- Modifying Perceptual Gender
- Rapid Prosody Transcription
- Pilot Study
- Discussion

Perceived Speaker Identity and Perception of Sociophonetic Variables

- Gender: Johnson, Strand, and D'Imperio (1999)
- Region of Origin: Niedzielski (1999), Hay and Drager (2010)
  - And Labov (1963) for speech production
- Sexual Orientation: Mack and Munson (2010)
- Socio-economic Status: Hay, Warren, and Drager (2006)
  - And Eckert (1989) for speech production
- Age: Koops, Gentry, and Pantos (2008)

Implicit or Explicit?

- No evidence of implicit use of social information:
  - Sexual Orientation: Mack and Munson (2010)
- Evidence of implicit use of social information:
  - Gender: Johnson, Strand, and D'Imperio (1999)
  - Region of Origin: Hay and Drager (2010)
  - Age: Koops, Gentry, and Pantos (2008)

What sociophonetic variables can be influenced?

- Vowels:
  - Johnson, Strand, and D'Imperio (1999)
  - Niedzielski (1999)
  - Hay and Drager (2010)
  - Hay, Warren, and Drager (2006)
  - Koops, Gentry, and Pantos (2008)
- Consonants:
  - Mack and Munson (2010)
  - Strand and Johnson (1996)

What sociophonetic variables can be influenced?

- What about suprasegmental components?
  - Prosodic prominence?

Prosodic Prominence in English

- Words that are perceived as “prominent” are typically produced acoustically louder, longer, and bear a pitch accent, relative to words that are not prominent. (Cole et al 2008)
  - What causes words to be louder, longer, or carry a pitch accent?

Prosodic Prominence in English

- Each phrase contains at least one accent. If there are more, there is one particularly salient pitch accent, known as the nuclear accent.
  - Typically appears phrase finally.
- Focus Projection (Selkirk 1995); Focus-to-Accent (Ladd 1996)
  - Accents can appear on words to due to their information status (new to the discourse, contrastive, etc)

Prosodic Prominence in English

- Word predictability correlates with duration, intensity and f0 with intended meaning (Watson 2009)
  - But these things tend to co-occur
- Listeners use their expectations of where prominence should fall (Wagner (2005), Cole, Mo, Hasegawa-Johnson (2008))
- Listeners also use visual information to perceive prominence (Krahmer and Swerts 2007)

Non-Social Factors Influencing Prominence Production

- The emotional state of the speaker (Pell (2000))
- The information status of the word and other words in the utterance (Breen et al (2010))
- The accentability of a word, such as function word (Calhoun 2010)
- Location within an utterance (Breen et al (2010), Xu and Xu (2005), Cooper et al (1985))

Non-Social Factors Influencing Prominence Production

- Physical gesturing by the speaker (Krahmer and Swerts (2007))
- The presence of another person (Breen et al (2010); Bux-Lugo et al (2013))
- The level of engagement of other speakers (Rosa et al 2012)
- The presence of ambiguity (Snedeker and Trueswell (2002))

Prominence Perception and Perceived Gender

- Gussenhoven and Rietveld (1998)
  - Effect of gender on degree of prominence in speech synthesized for gender
  - But:
    - Is masculine speech perceived as more prominent than feminine speech (scaling)?
    - Or are prosodic events more likely to be perceived in male speech than female speech (linguistic)?

Research Questions

- In what way is the perception of prominence in English influenced by the perception of speaker gender?
  - Discrete or gradient?
  - Implicit or explicit association?

- To investigate these, we'll need:
  - Resynthesized speech
  - A method for collecting prosodic scores/annotations

Overview

- Introduction
- Modifying Perceptual Gender
- Rapid Prosody Transcription
- Pilot Study

Resynthesizing for Gender

- Need to scale the formants while leaving other aspects of the speech alone.

Praat KlattGrid (Weenink 2009)

- Implementation of the Klaat Synthesizer (Klaat and Klaat 1990)
- Discrete parameterization of the speechwave
  - Powerful but complicated
- Workflow:
  - Extract KlattGrid parameters from source audio file
  - Conduct manipulation on KlattGrid
  - Resynthesize source audio with modified KlattGrid

Praat KlattGrid (Weenink 2009)

- My python interface to KlattGrid (praatIO) is available as a freely available resource
  - Allows for programmable manipulation of klattgrids.
  - https://github.com/timmahrt/praatIO

Original Audio

Formants increased 20%

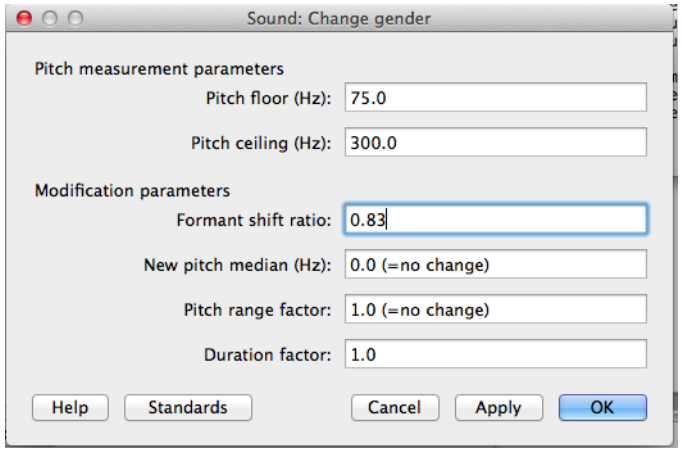
Formants Decreased 20%

Other Resynthesis Tools

- TANDEM-STRAIGHT (Kawahara et al 2008)
  - Powerful and feature rich, but also more complicated
- Audiosculpt (Centre Pompidou)

Praat “Change Gender”

- Easy to use; high-quality results.
- Formants are shifted by “manipulation of the sampling frequency”



Praat “Change Gender”

- Very good results can be obtained by manipulating both formants and mean pitch:



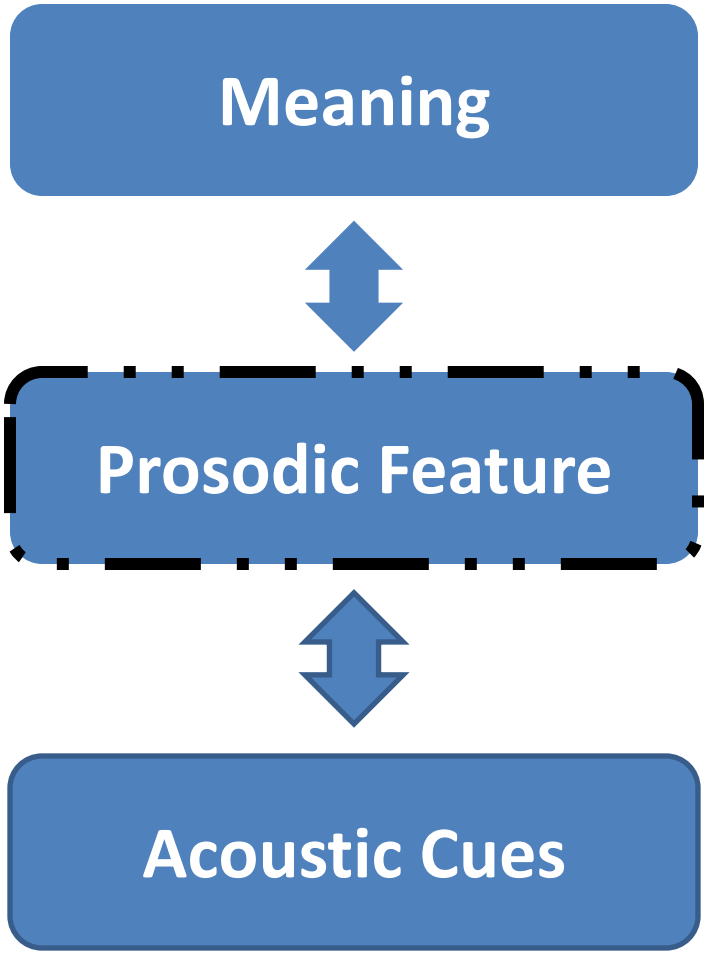

Overview

- Introduction
- Modifying Perceptual Gender
- Rapid Prosody Transcription
- Pilot Study

Prosodic annotation is (still) difficult: WHY?

- Cue weighting, Reduced or ambiguous cues
- Individual speaker differences
- Context effects (syntax, discourse); top-down processing
- Disfluency
- Uncertain contrast among labels

These factors are especially challenging in spontaneous speech, where communicative functions of prosody and phonetic reduction patterns are highly variable.



Disagreeing annotators

- Variable and ambiguous acoustic cues to prosody can yield different prosodic labels from different annotations of the same utterance.
- Disagreement = Noise
- Disagreements resolved through consensus, majority vote, or arbitration to produce a final annotation.

Noise = information

- Inter-annotator disagreement on a prosodic label can be informative, revealing ambiguities and other issues in need of further research.
- Multiple independent expert annotators as a standard?
- Problem: time, training, money

Crowdsourcing: “Labels for less”

Hasegawa-Johnson et al. (2015):

- Factor a difficult annotation task into a set of binary-coded “easy” questions that can be answered by anyone.
- Use crowd-sourcing, trade accuracy and specificity of annotations for efficiency and lower cost
- Scientific benefits: better modeling of form-function mapping are possible with larger labeled datasets

Crowdsourcing prosodic annotation

Rapid Prosodic Transcription (RPT), using

- Coarse-grained labels: Prominence, Boundary
- Real-time annotation
- Non-expert, untrained annotators
- Based on auditory impression- no visual speech display

RPT method

The annotator listens to speech sample of up to 1 minute duration, and follows along on a printed transcript with no punctuation or capitalization.

Round 1: boundaries	Select the location of boundaries that separate perceived chunks of speech:
	…it's gonna happen | that our society…

RPT method

The annotator listens to speech sample of up to 1 minute duration, and follows along on a printed transcript with no punctuation or capitalization.

Round 1: boundaries	Select the location of boundaries that separate perceived chunks of speech:
	…it's gonna happen | that our society…
Round 2: prominences	Selects words perceived as prominent:
	…it's gonna **happen** that **our** society…

Coding and pooling data

- Each annotator's annotations are coded on each word as a 0 or 1 for Prominence, and 0 or 1 for Boundary.
- Annotations are pooled over annotators

TOKEN	WORD
1	i
2	really
3	don't
4	know
5	i
6	think
7	in
8	today's
9	world
10	what
11	they
12	call
13	the
14	nineties
15	that
16	uh
17	it's
19	like

TOKEN	WORD
1	i
2	really
3	don't
4	know
5	i
6	think
7	in
8	today's
9	world
10	what
11	they
12	call
13	the
14	nineties
15	that
16	uh
17	it's
19	like

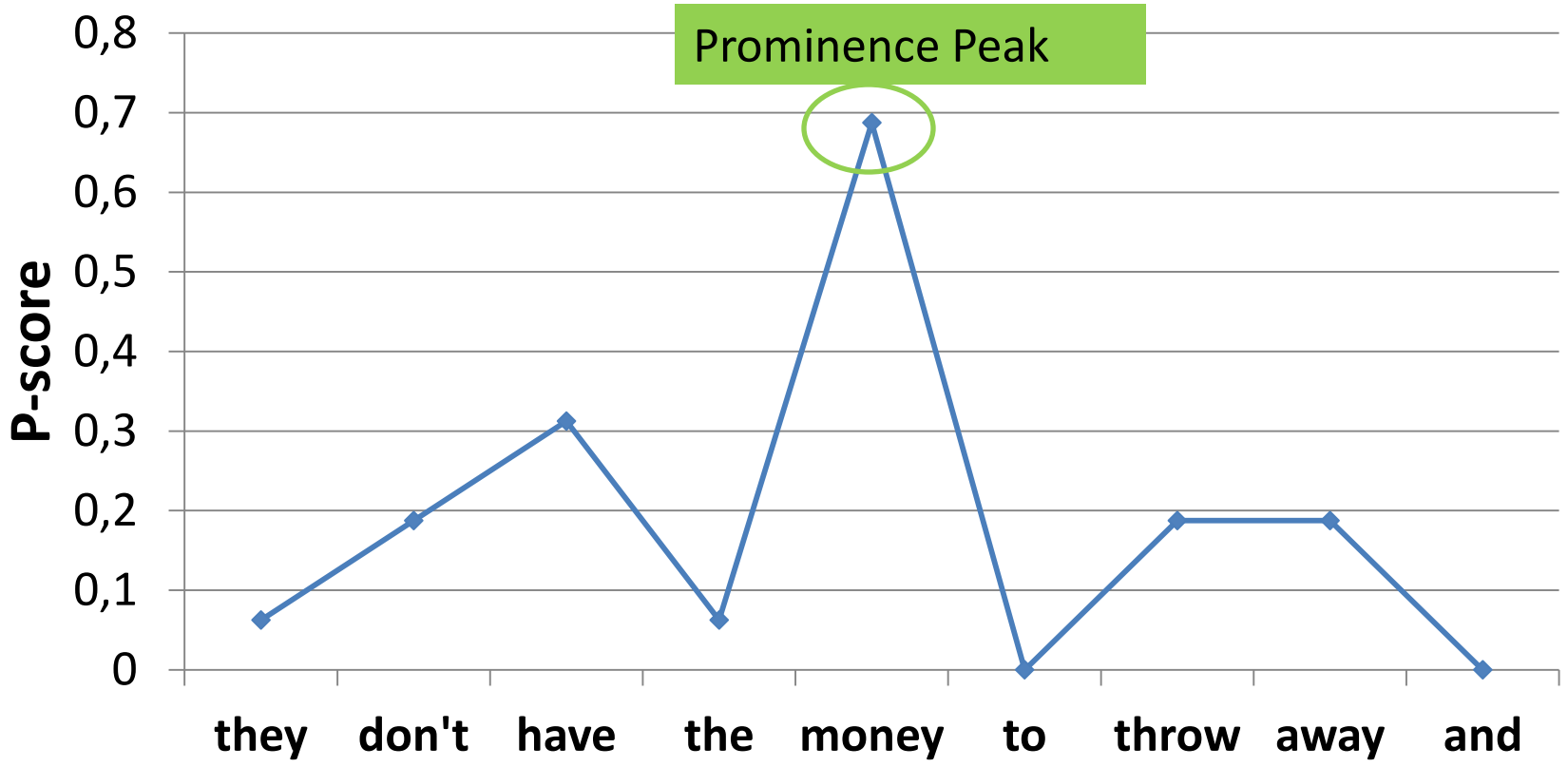
Boundary				Prominence				
A1	A2	A3	A4	A1	A2	A3	A4	…
0	0	0	0	0	0	0	0	
0	0	0	0	0	1	1	0	
0	0	0	0	0	0	0	0	
1	1	1	0	0	0	0	0	
0	0	0	0	0	0	0	0	
0	0	0	0	0	1	0	0	
0	0	0	0	0	0	0	0	
0	0	0	0	1	1	0	0	
1	1	1	0	0	0	0	0	
0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	
1	1	1	0	1	0	0	1	
0	1	1	0	0	0	0	0	
1	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	
0	0	0	0	0	0	0	0	

1 annotator's labels

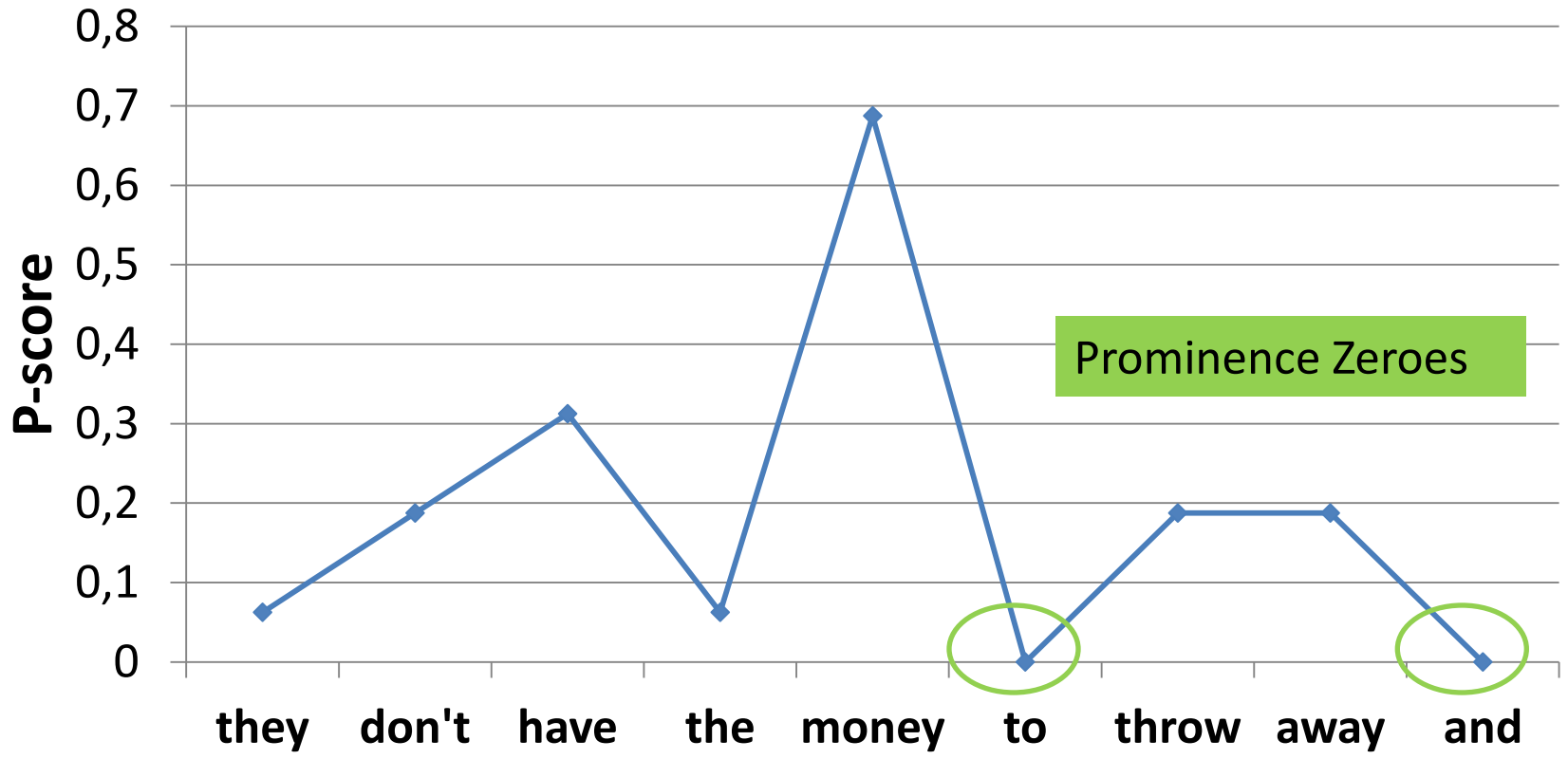
				Boundary				Prominence				
TOKEN	WORD	B-SCORE	P-SCORE	A1	A2	A3	A4	A1	A2	A3	A4	…
1	i	0	0	0	0	0	0	0	0	0	0	
2	really	0	0.56	0	0	0	0	0	1	1	0	
3	don't	0	0.25	0	0	0	0	0	0	0	0	
4	know	0.81	0.44	1	1	1	0	0	0	0	0	
5	i	0	0	0	0	0	0	0	0	0	0	
6	think	0.31	0.44	0	0	0	0	0	1	0	0	
7	in	0	0	0	0	0	0	0	0	0	0	
8	today's	0	0.81	0	0	0	0	1	1	0	0	
9	world	0.81	0.44	1	1	1	0	0	0	0	0	
10	what	0	0	0	0	0	0	0	0	0	0	
11	they	0	0	0	0	0	0	0	0	0	0	
12	call	0	0	0	0	0	0	0	0	0	0	
13	the	0	0	0	0	0	0	0	0	0	0	
14	nineties	0.81	0.5	1	1	1	0	1	0	0	1	
15	that	0.5	0.13	0	1	1	0	0	0	0	0	
16	uh	0.38	0	1	0	0	0	0	0	0	0	
17	it's	0	0	0	0	0	0	0	0	0	0	
19	like	0	0.06	0	0	0	0	0	0	0	0	

Prominences and Boundaries coded as counts or proportions

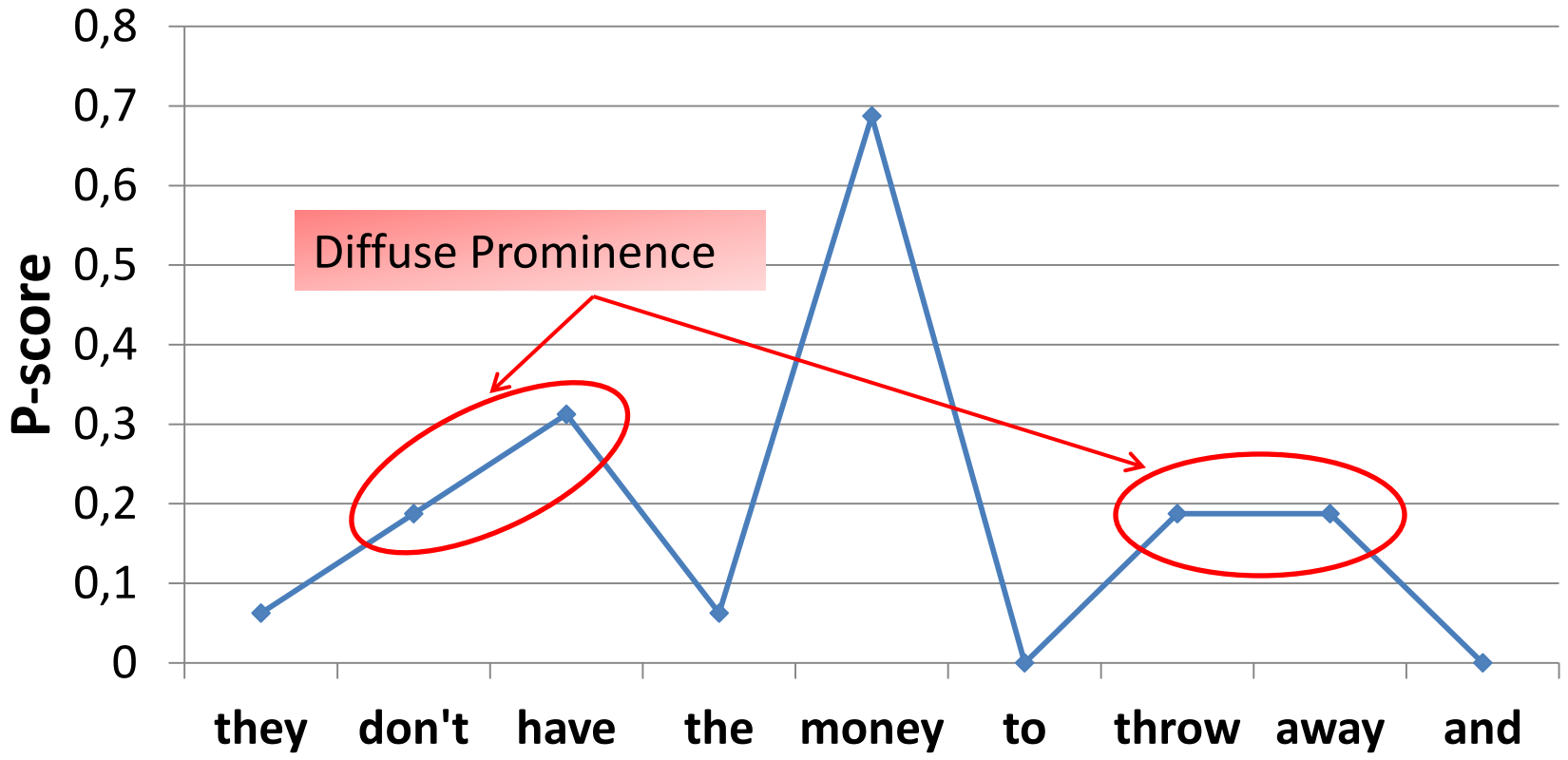
Prominence scores (proportions) plotted for each word in an (partial) utterance.



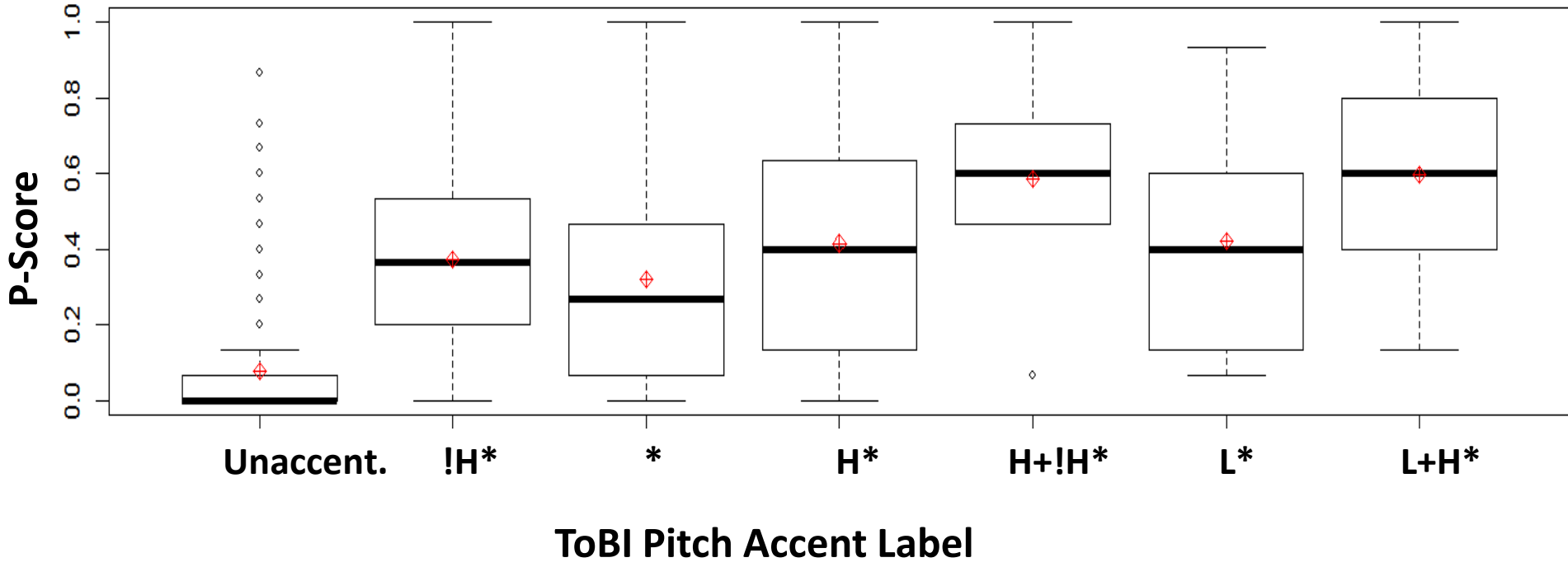
Prominence scores (proportions) plotted for each word in an (partial) utterance.



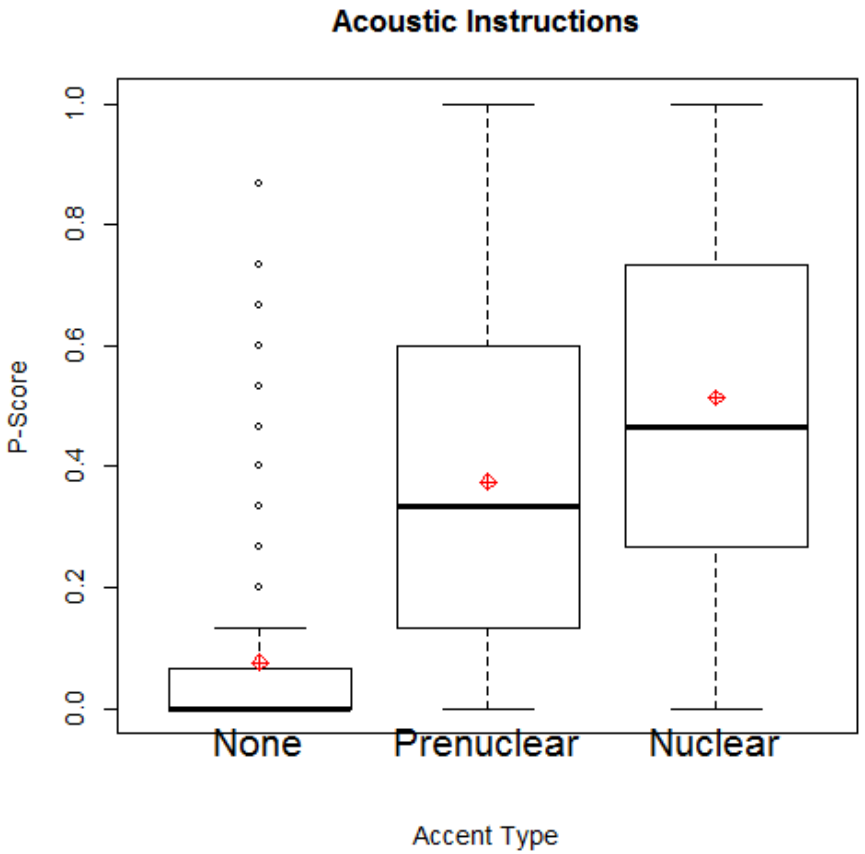
Prominence scores (proportions) plotted for each word in an (partial) utterance.



Prominence ratings by ToBI pitch accent type (American English)



Prominence ratings by ToBI pitch accent type (American English)



Overview

- Introduction
- Modifying Perceptual Gender
- Rapid Prosody Transcription
- Pilot Study

Research Question

- In what way is the perception of prominence in English influenced by the perception of speaker gender?
  - Hypotheses?

Stimuli

- 5 utterances produced by males, where the formant resynthesis led to a different interpretation of their gender
  - 19-27 seconds long
  - 365 words total
- Taken from:
  - Buckeye Corpus (Pitt et al 2007)
  - CHAINS Corpus (Cummins et al 2006)
  - SBCSAE Corpus (DuBois et al 2000-2005)

Stimuli

- 15 utterances
  - 5 original utterances
  - 5 with formants increased 20%
  - 5 with formants decreased 17%
- Resynthesis performed using Praat's (Boersma and Weenink 2015) "Change Gender" function

Stimuli

Original

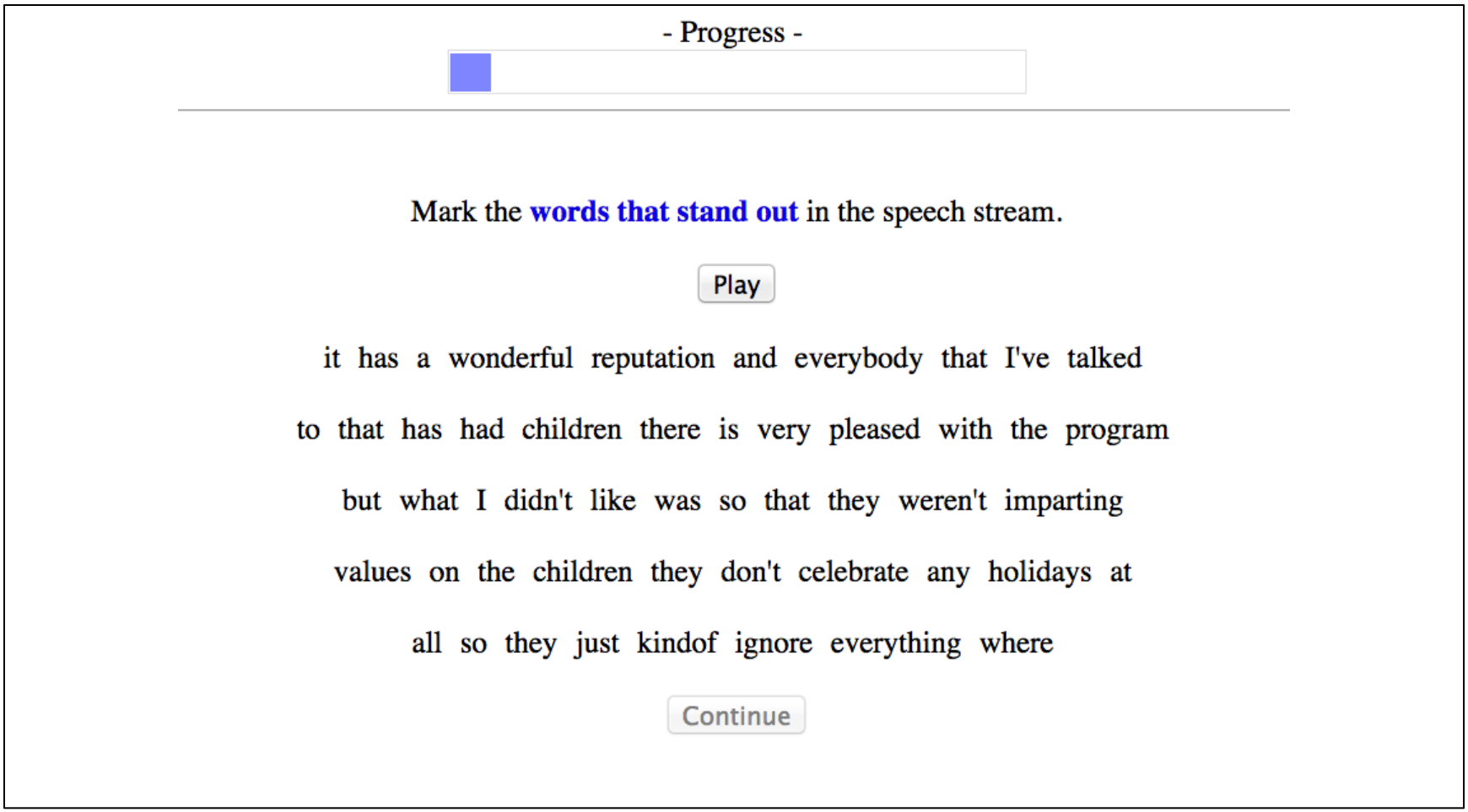
Increased formants

Decreased formants

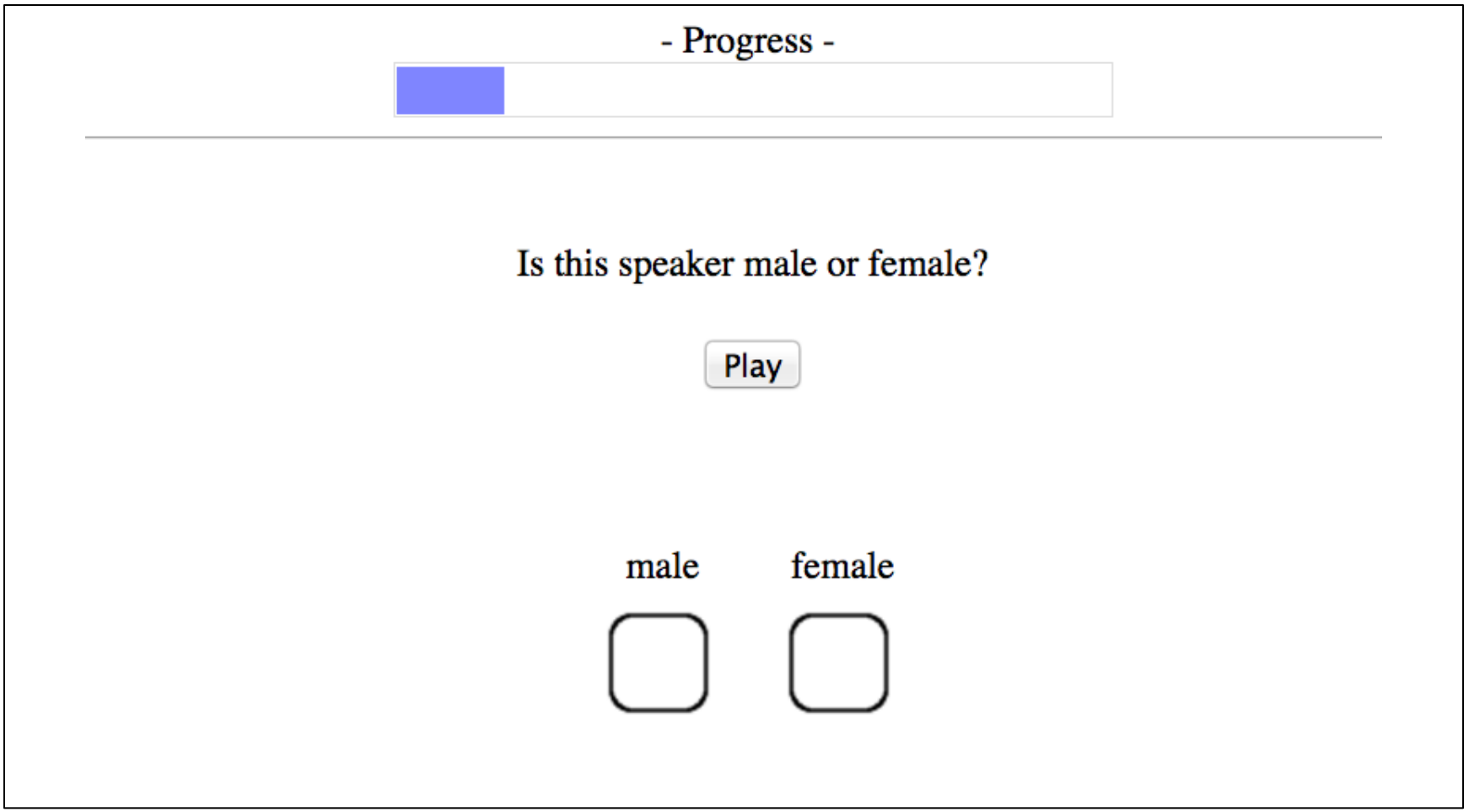
Participants

- 3 Groups
  - Original utterances
  - Increased formants
  - Decreased formants
- Recruited over Amazon Mechanical Turk
  - IP address limited to the United States
- Experiment administered over LMEDS (Mahrt 2015)
  - Download at: https://github.com/timmahrt/LMEDS

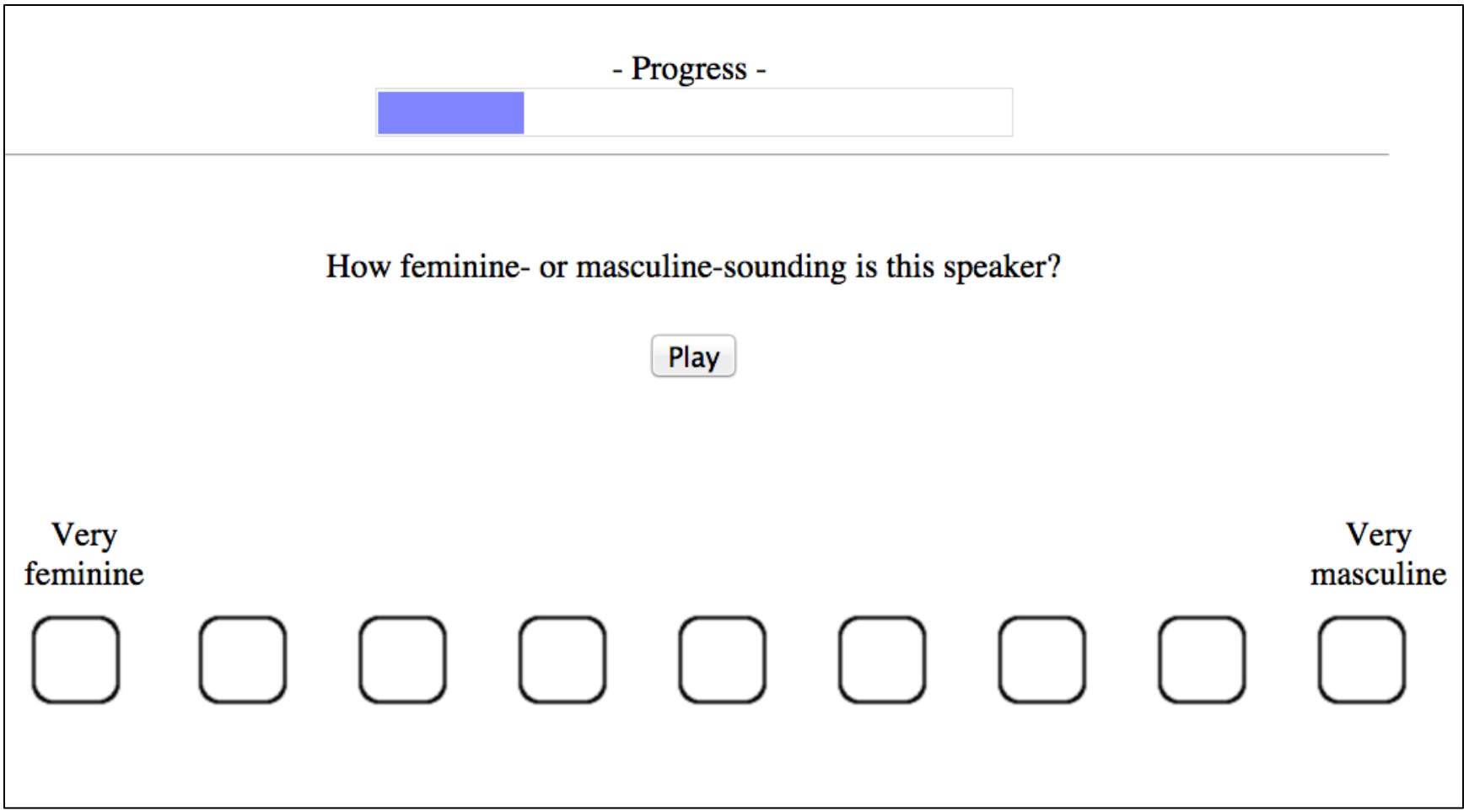
Task 1

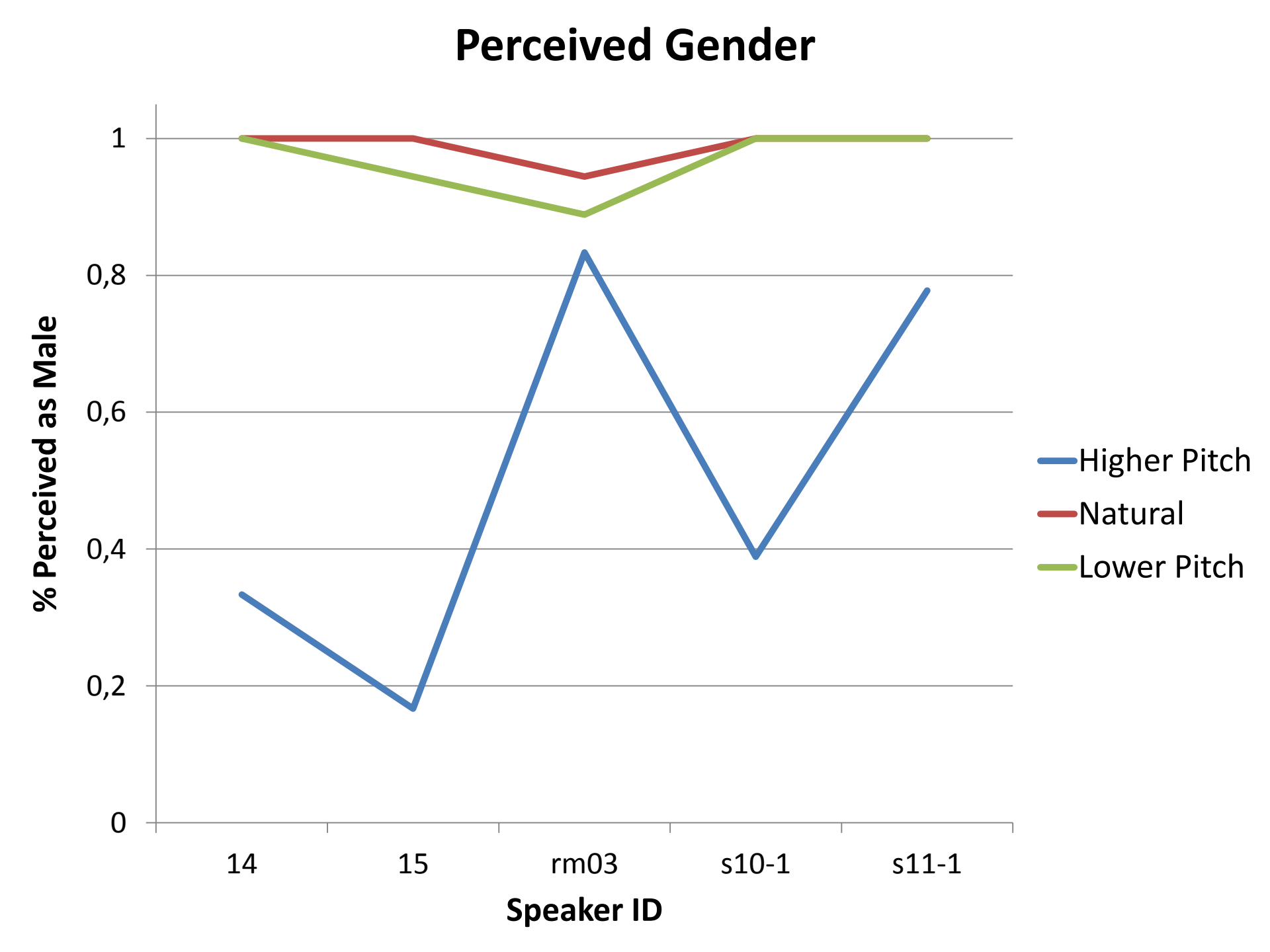

- Progress -
Mark the words that stand out in the speech stream.
Play
it has a wonderful reputation and everybody that I've talked
to that has had children there is very pleased with the program
but what I didn't like was so that they weren't imparting
values on the children they don't celebrate any holidays at
all so they just kindof ignore everything where
Continue

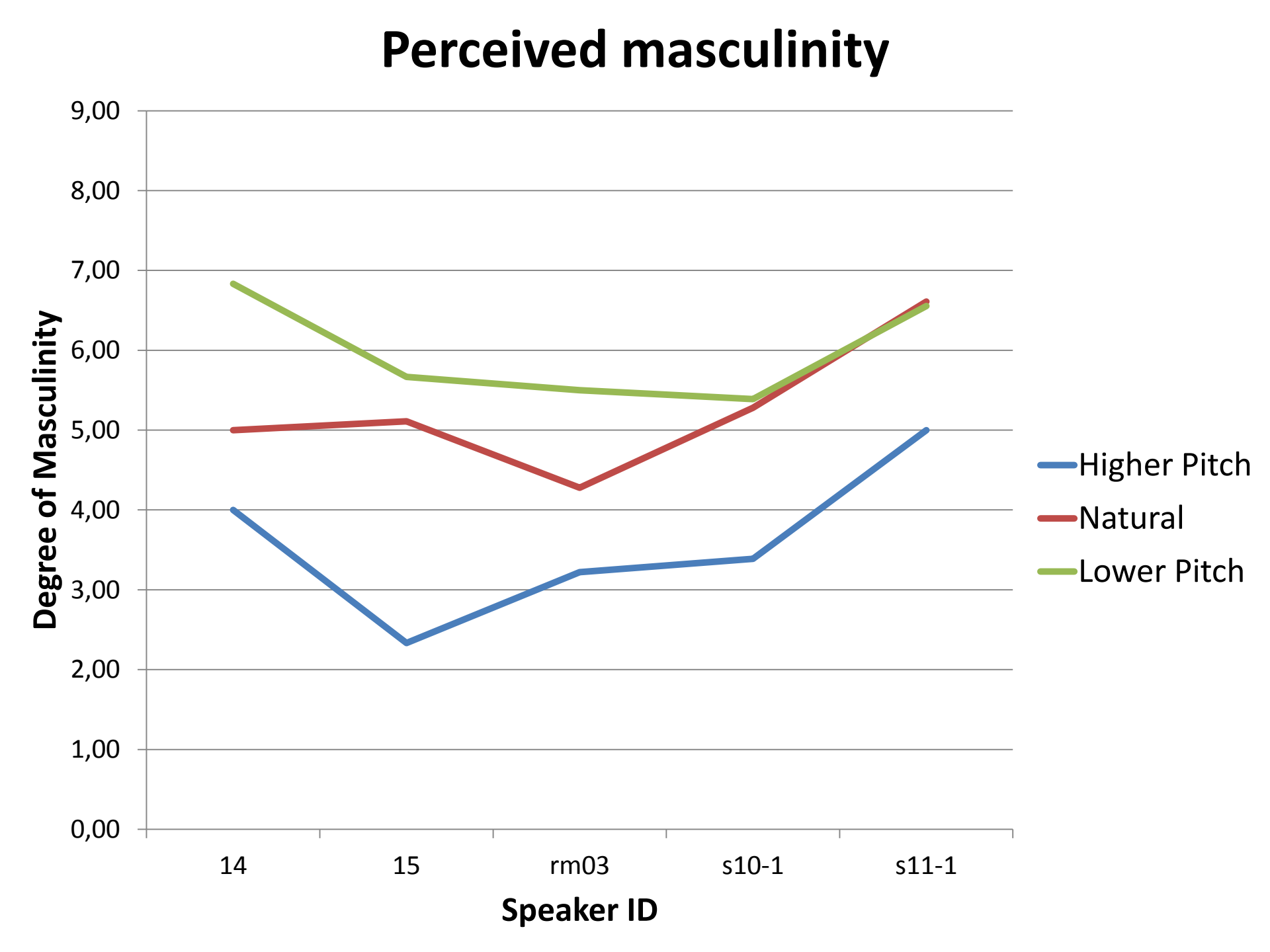
Task 2

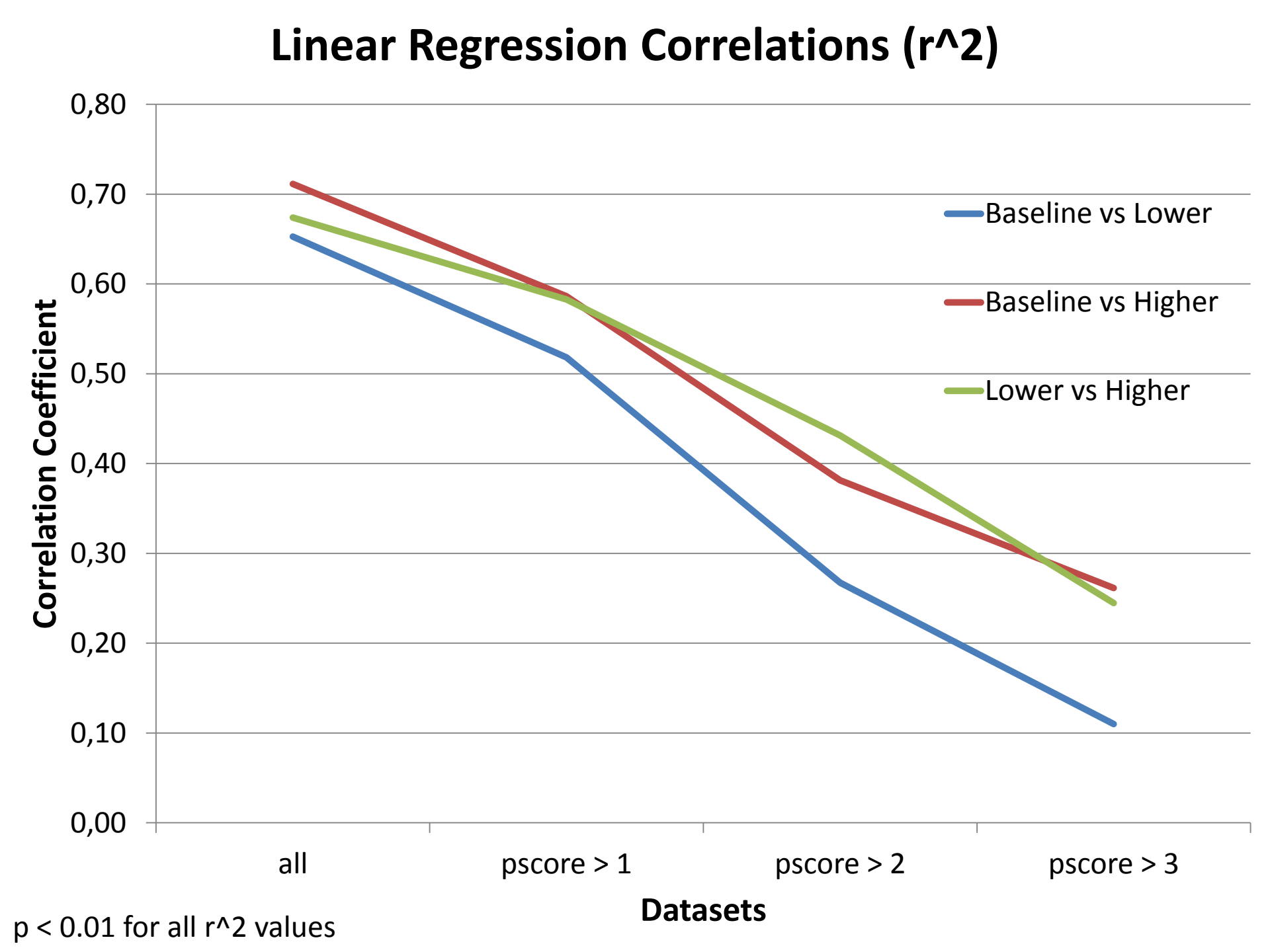


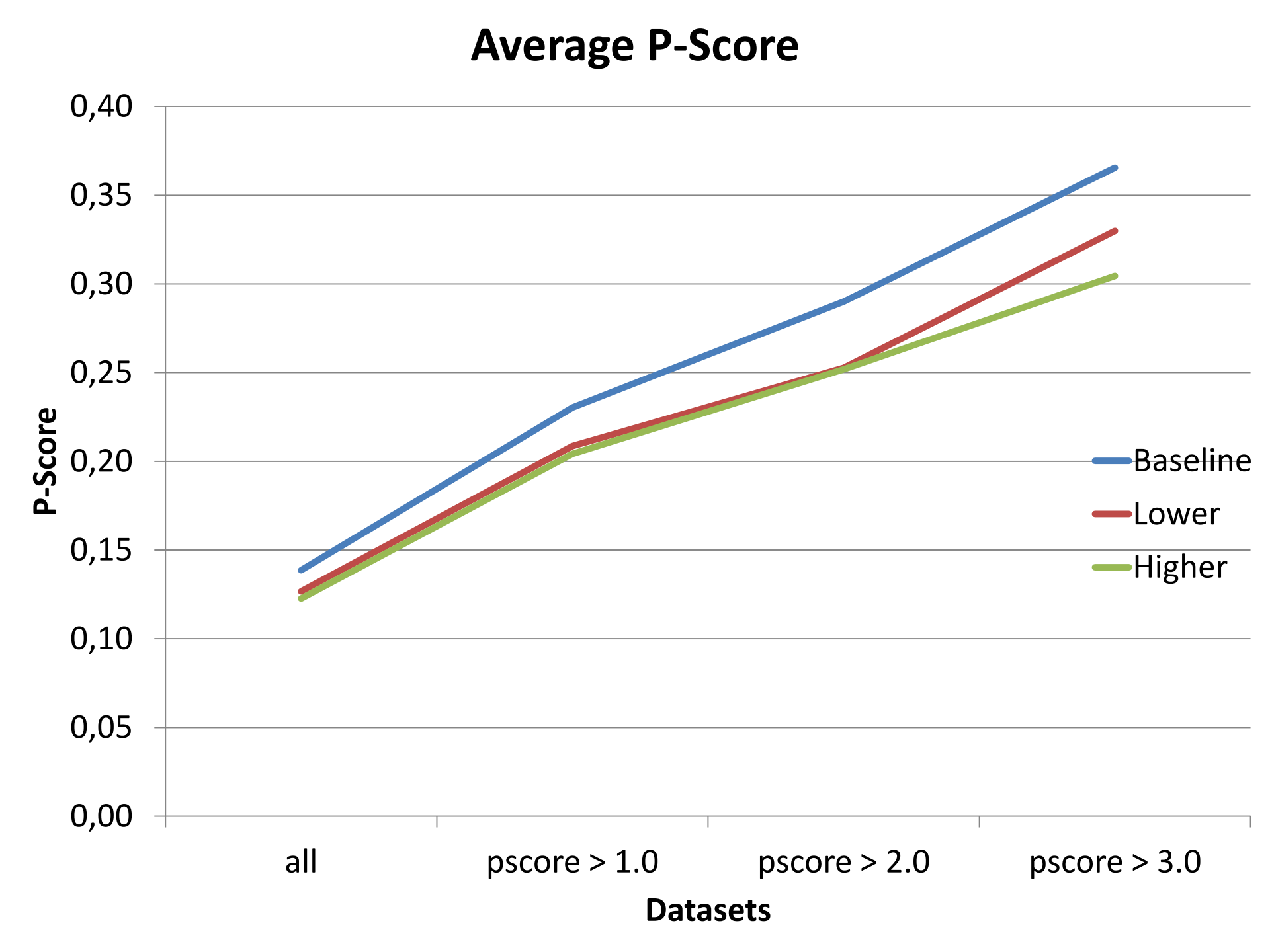
Task 3


- Progress -
How feminine- or masculine-sounding is this speaker?
Play
Very feminine
Very masculine


Perceived Gender
% Perceived as Male
1
0,8
0,6
0,4
0,2
0
14
15
rm03
s10-1
s11-1
Speaker ID
Higher Pitch
Natural
Lower Pitch


Perceived masculinity
9,00
8,00
7,00
6,00
5,00
4,00
3,00
2,00
1,00
0,00
Degree of Masculinity
14
15
rm03
s10-1
s11-1
Speaker ID
Higher Pitch
Natural
Lower Pitch


Linear Regression Correlations (r^2)
0,80
0,70
0,60
0,50
0,40
0,30
0,20
0,10
0,00
Correlation Coefficient
Baseline vs Lower
Baseline vs Higher
Lower vs Higher
all
pscore > 1
pscore > 2
pscore > 3
Datasets
p < 0.01 for all r^2 values


Average P-Score
0,40
0,35
0,30
0,25
0,20
0,15
0,10
0,05
0,00
P-Score
Baseline
Lower
Higher
all
pscore > 1.0
pscore > 2.0
pscore > 3.0
Datasets

Findings

- Increased formants lead to a greater chance of a man being perceived as female. (But only for certain speakers (why?))
- Formant height may be inversely correlated with masculinity
- Increased formants p-scores are more similar to the baseline than the decreased formants p-scores. But p-scores are higher for the baseline than the manipulated results.

Discussion

- Why a different result from Gussenhoven and Rietveld 1998?
  - They used male and female speakers with only resynthesized pitch.
  - Perhaps this lends evidence to the idea that although there are differences in the degree of prominence—prominent words are still perceived as prominent.

Developing the Study Further

- Add more speakers—more masculine male speakers and feminine and masculine female speakers
- New task—repetition of Gussenhoven and Rietveld 1998.
- New stimuli with pitch and formants manipulated?

Comments?
Questions?

Thank you for listening