

Documenting Perception

Survey, Transcription, Annotation

Dafydd Gibbon

Guangzhou Prosody Lectures 2017

Documenting perception: Overview

- Three phases of speech: production, transmission, perception
 - Perception – transcription - annotation
 - From perception to transcription, from transcription to annotation
 - Speech perception
- Another kind of documentation of perception: opinion survey
 - Likert Scale Opinion Survey, survey Data Flow
 - Questionnaire design quantitative summary, evaluation
 - Likert scores for all submissions for each attribute
 - Average value for each attribute for each item
 - Visualisations of average attribute scores, Tones 1, 2, 3, 4
- Free interview survey format: selection of comments
- Conclusions

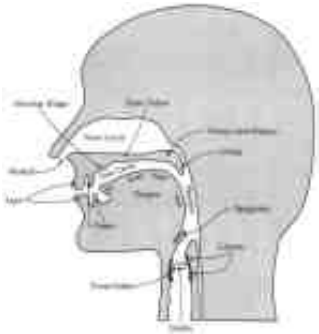
Three phases of speech: production, transmission, perception



A tiger and a mouse were walking in a field...

Three phases of speech: production, transmission, perception

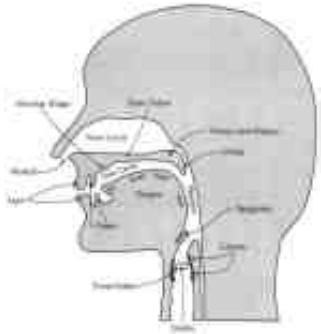
Articulatory Phonetics



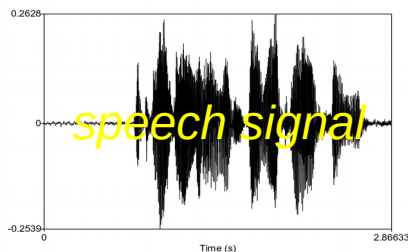
A tiger and a mouse were walking in a field...

Three phases of speech: production, transmission, perception

Articulatory
Phonetics



A tiger and a mouse were walking in a field...



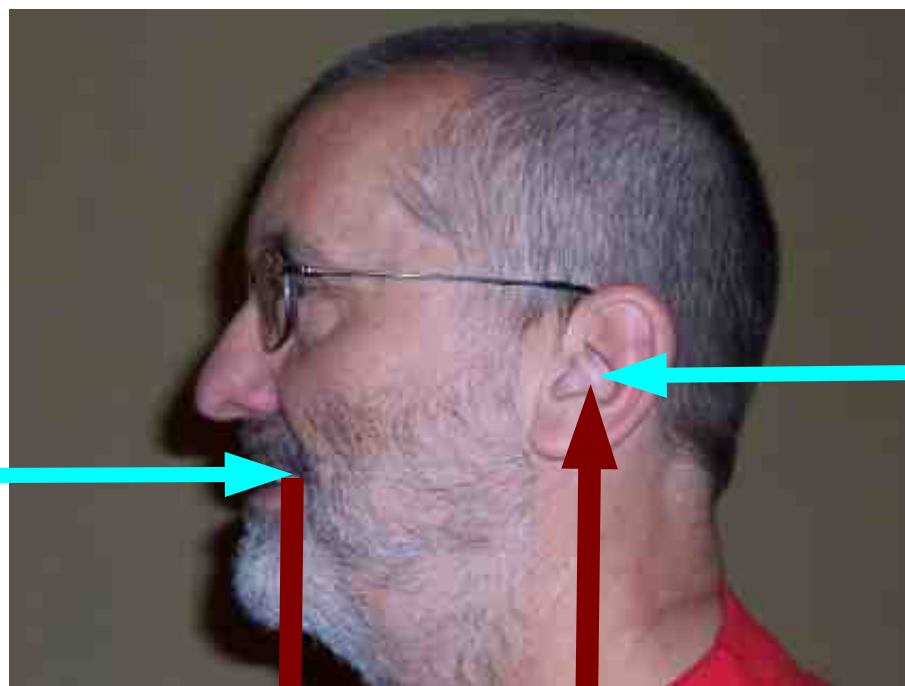
Acoustic
Phonetics

Three phases of speech: production, transmission, perception

Articulatory
Phonetics

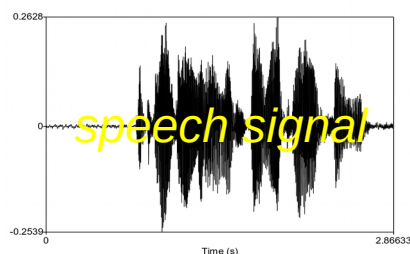


Auditory
Phonetics



A tiger and a mouse were walking in a field...

Acoustic
Phonetics

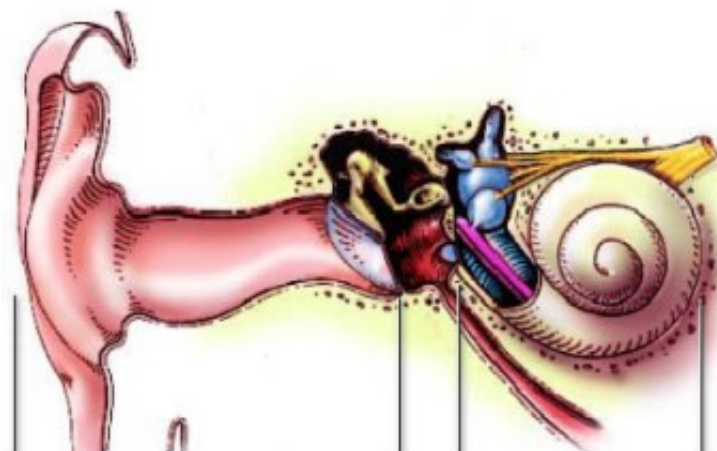


Perception – transcription - annotation

- Auditory perception underlies transcription
 - continuous perception
 - categorial perception
- Transcription is one kind of
 - documentation of categorial perception
 - usually using the International Phonetic Alphabet
 - or specialised alphabets
 - for prosody
 - for speech pathology
- Annotation is the assignment
 - of segments of a speech signal (segmentation)
 - to a transcription (classification)

From perception to transcription, from transcription to annotation

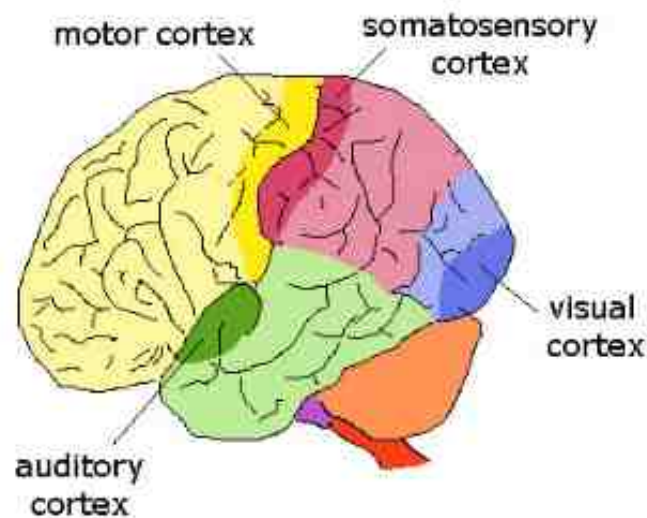
The Ear



outer ear

middle ear

inner ear

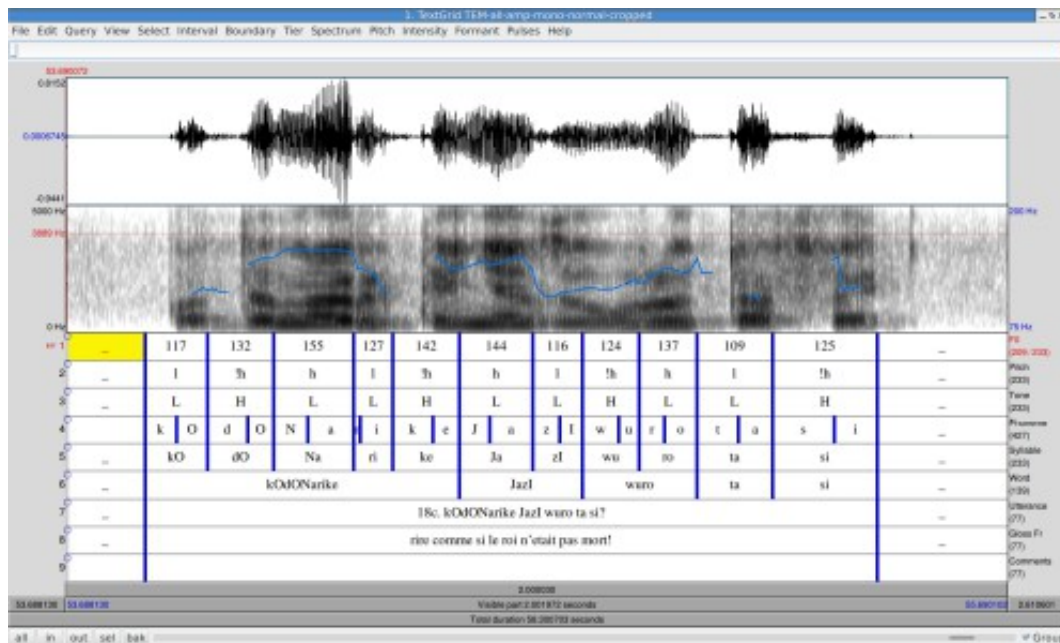


motor cortex

somatosensory cortex

visual cortex

auditory cortex



Underlying theories

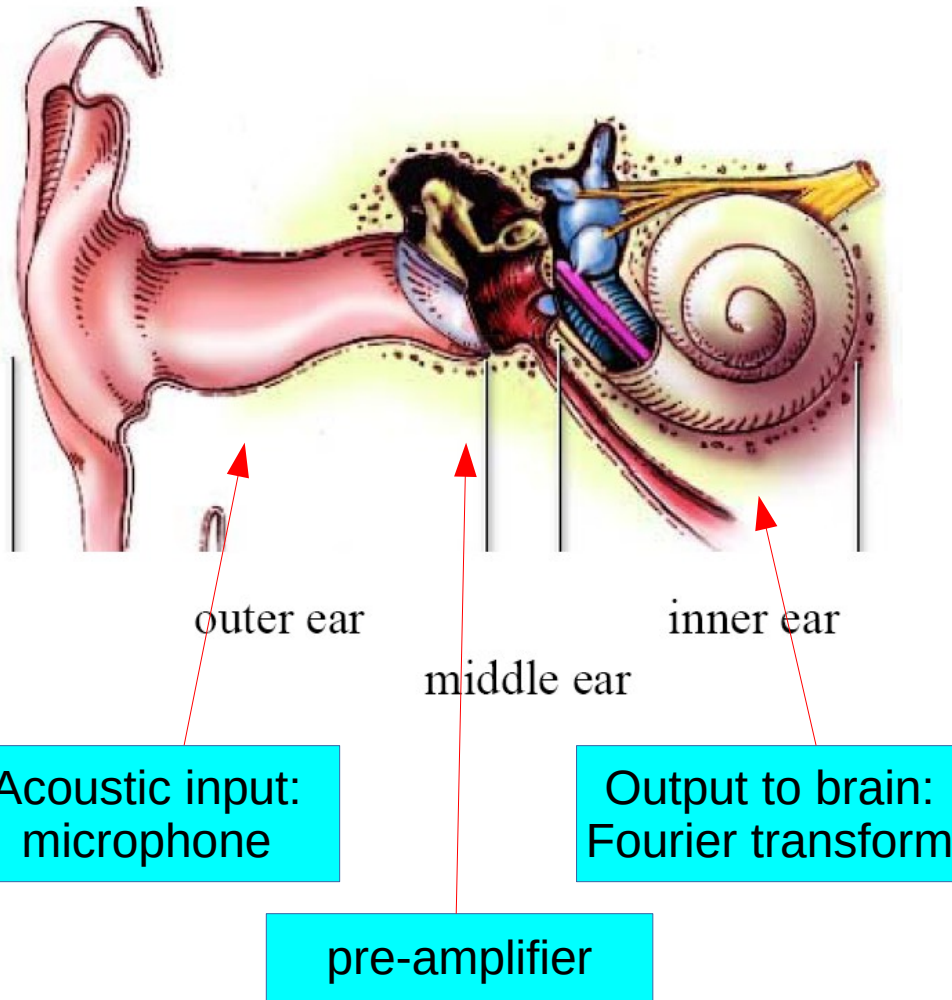
categorical perception
segmentation and classification

signal processing
graphical visualisation

computer software (e.g. Praat)

Speech perception

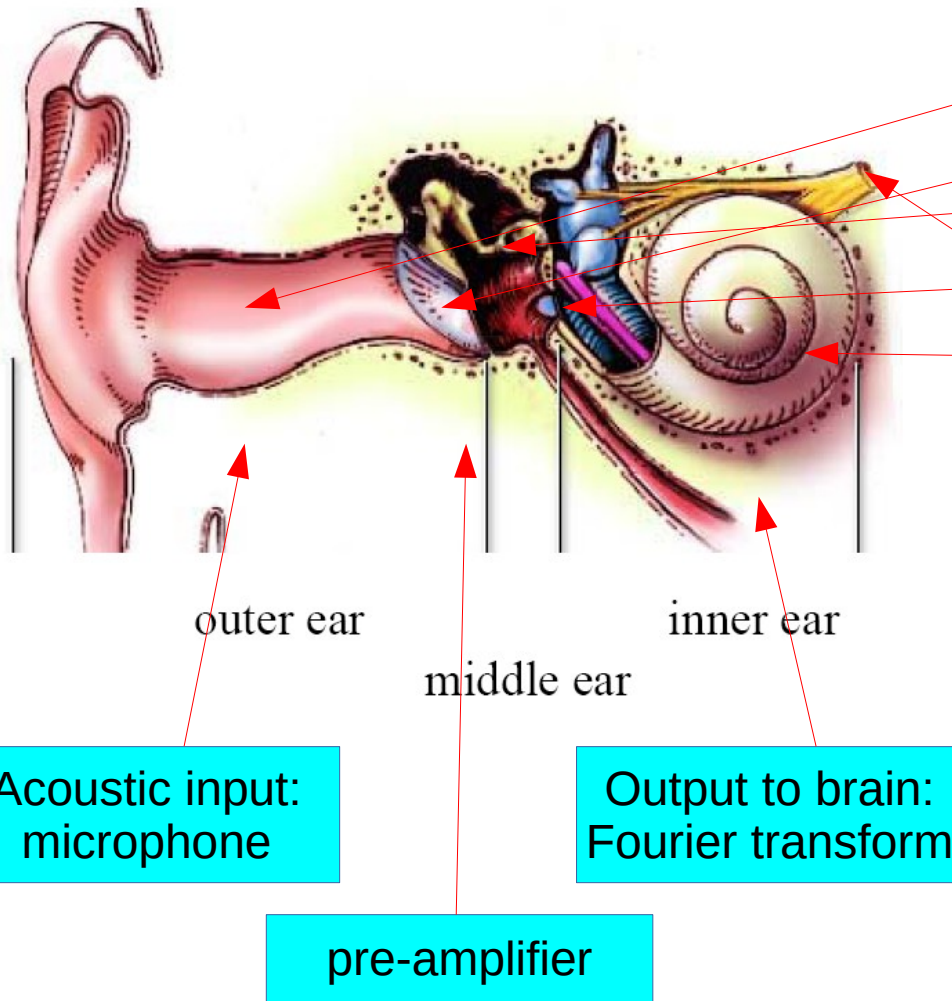
The Ear



- Anatomical and physiological phases:
 - auditory canal
 - eardrum
 - ossicles ('little bones')
 - oval window
 - cochlea ('snail')
 - auditory nerve
 - brain:
 - Left hemisphere
 - Right hemisphere
- Cognitive phases:
 - Short term memory
 - categorial perception
 - Long term memory

Speech perception

The Ear



- Anatomical and physiological phases:

- auditory canal
- eardrum
- ossicles ('little bones')
- oval window
- cochlea ('snail')
- auditory nerve
- brain:
 - Left hemisphere
 - Right hemisphere

- Cognitive phases:

- Short term memory
 - categorial perception
- Long term memory

Another kind of documentation of perception: opinion survey

- Metalinguistic documentation of perception:
 - Transcriptions
 - opinion surveys
- Indirect representation of perception
- Not direct measurements of perception
- More direct measurements are performed, for example, by
 - reaction time experiments
 - neurophysiological experiments

Likert Scale Opinion Survey

- There are many formats for opinion surveys
open interview ... closed set (standard: Likert scale)
- For this test:
 - audio input
 - Likert scale response:
agree ... disagree
with a statement

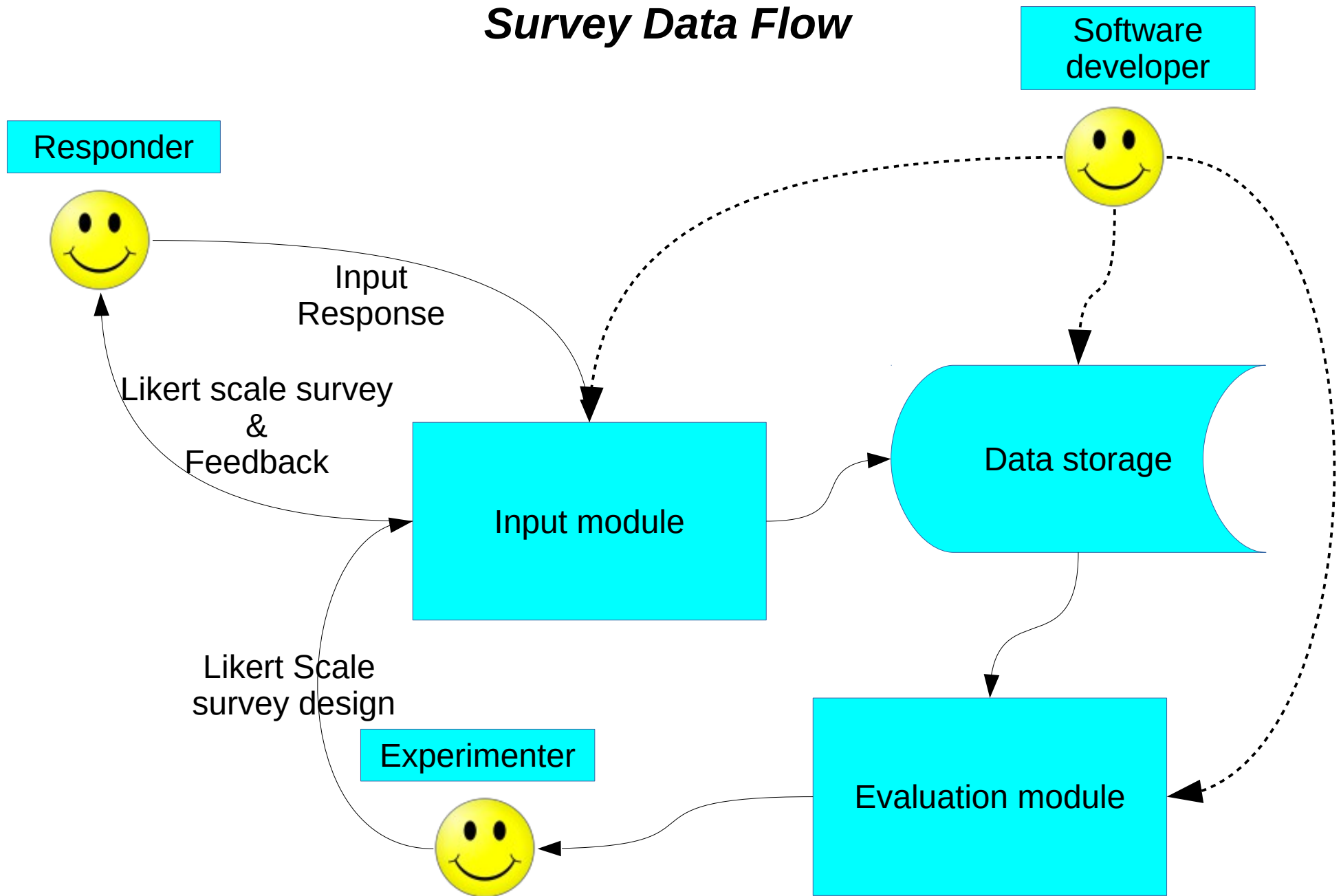


Dr. Rensis Likert
1903-1981

Likert scale 1932
(Ph.D. thesis)

<i>The melody of the sample is...</i>	yes	maybe	unsure	maybe not	no
<i>high</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>rising</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>mid</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>rising-falling</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>low</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>falling-rising</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>level</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>falling</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

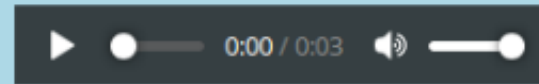
Survey Data Flow



Questionnaire design and quantitative summary

Audio_1:

Listen to the recording at least twice:



The melody of the sample is...

	yes	maybe	unsure	maybe not	no
<i>high</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>rising</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>mid</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>rising-falling</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>low</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>falling-rising</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>level</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
<i>falling</i>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Please comment on the pronunciation of this example. Is it normal?
Anything special?

Likert opinion scale options:	5
Attributes:	8
Comment fields:	1
Prompts (audio, transcripts):	16
Attribute instances per submission:	128
Comment instances per submission:	16
Submission count:	20

Questionnaire evaluation

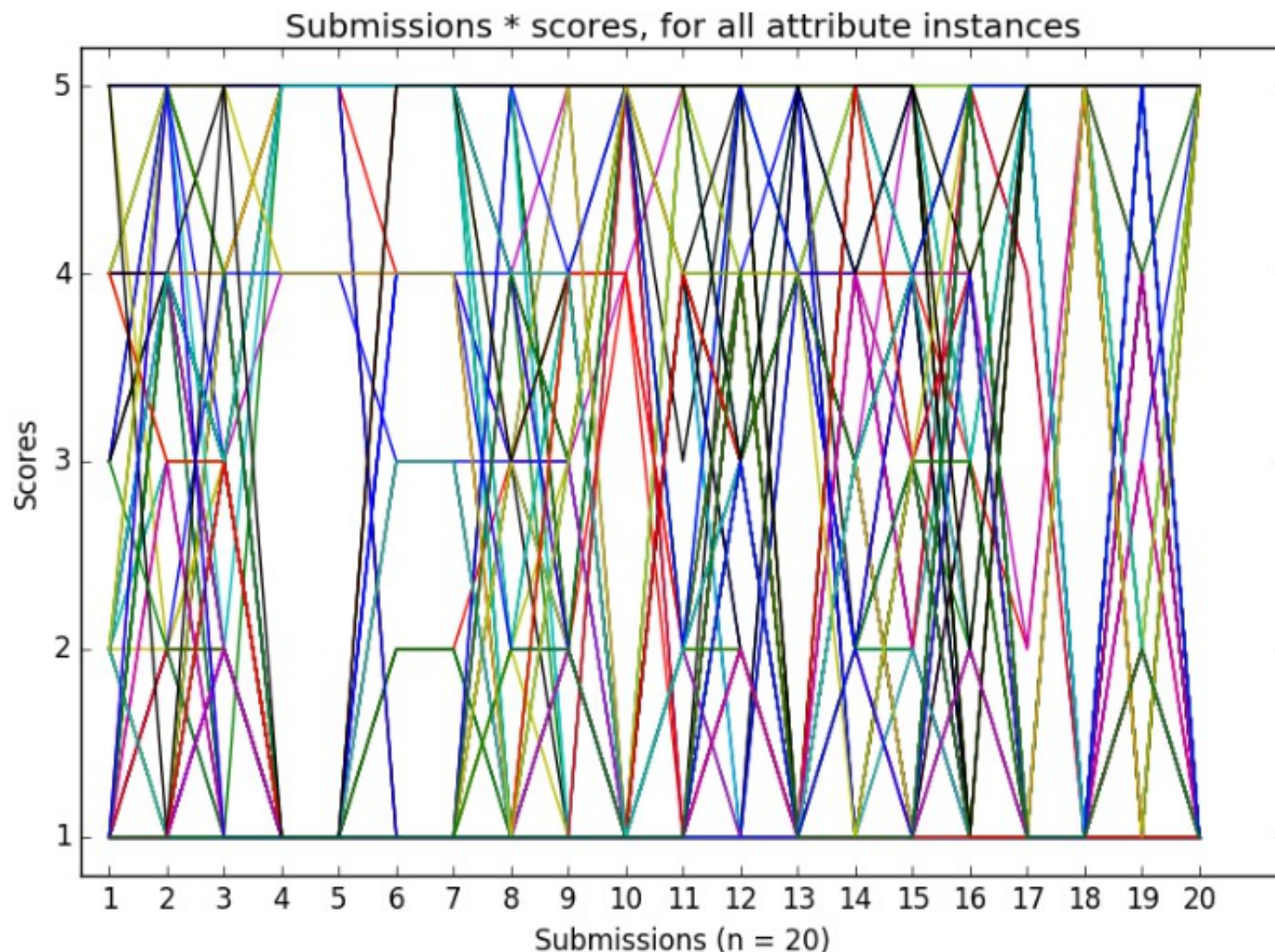
OSCAR implementation at Bielefeld University (D. Gibbon)

Tables can be copied and pasted into a text editor and saved as text file with extension '.csv'.

The CSV file can be loaded into a spreadsheet for further processing.

Figures can be saved and used in documents.

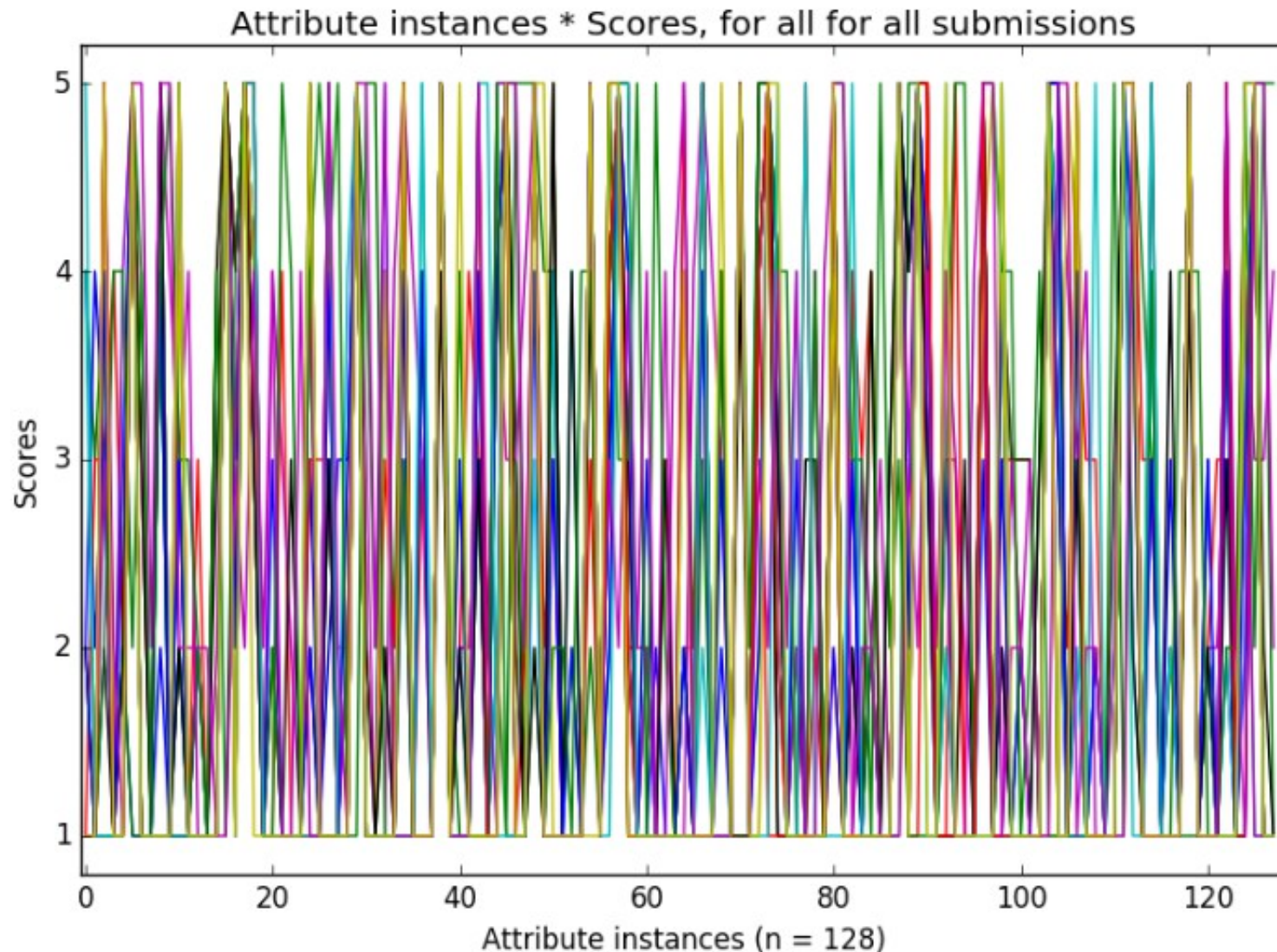
Likert scores for all submissions for each attribute



Each line represents the Likert scores for one of the 8 attributes:
high, rising, mid, rising-falling, low, falling-rising, level, falling

(Not very useful unless the responders tend to agree.)

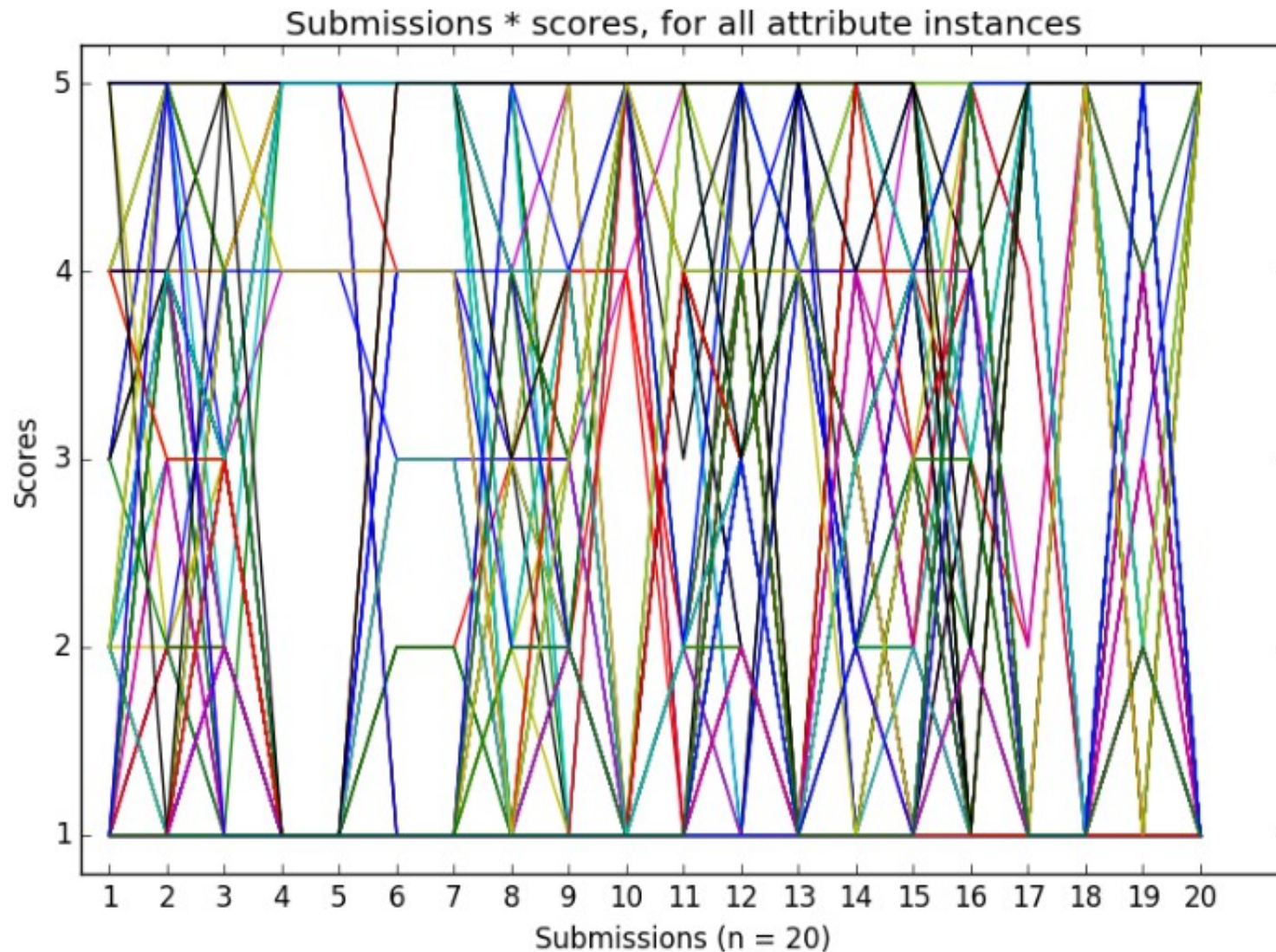
Likert scores for each attribute/item for all submissions



Each line represents the Likert scores for one of the 18 submissions:
high, rising, mid, rising-falling, low, falling-rising, level, falling

(Not very useful unless the responders tend to agree.)

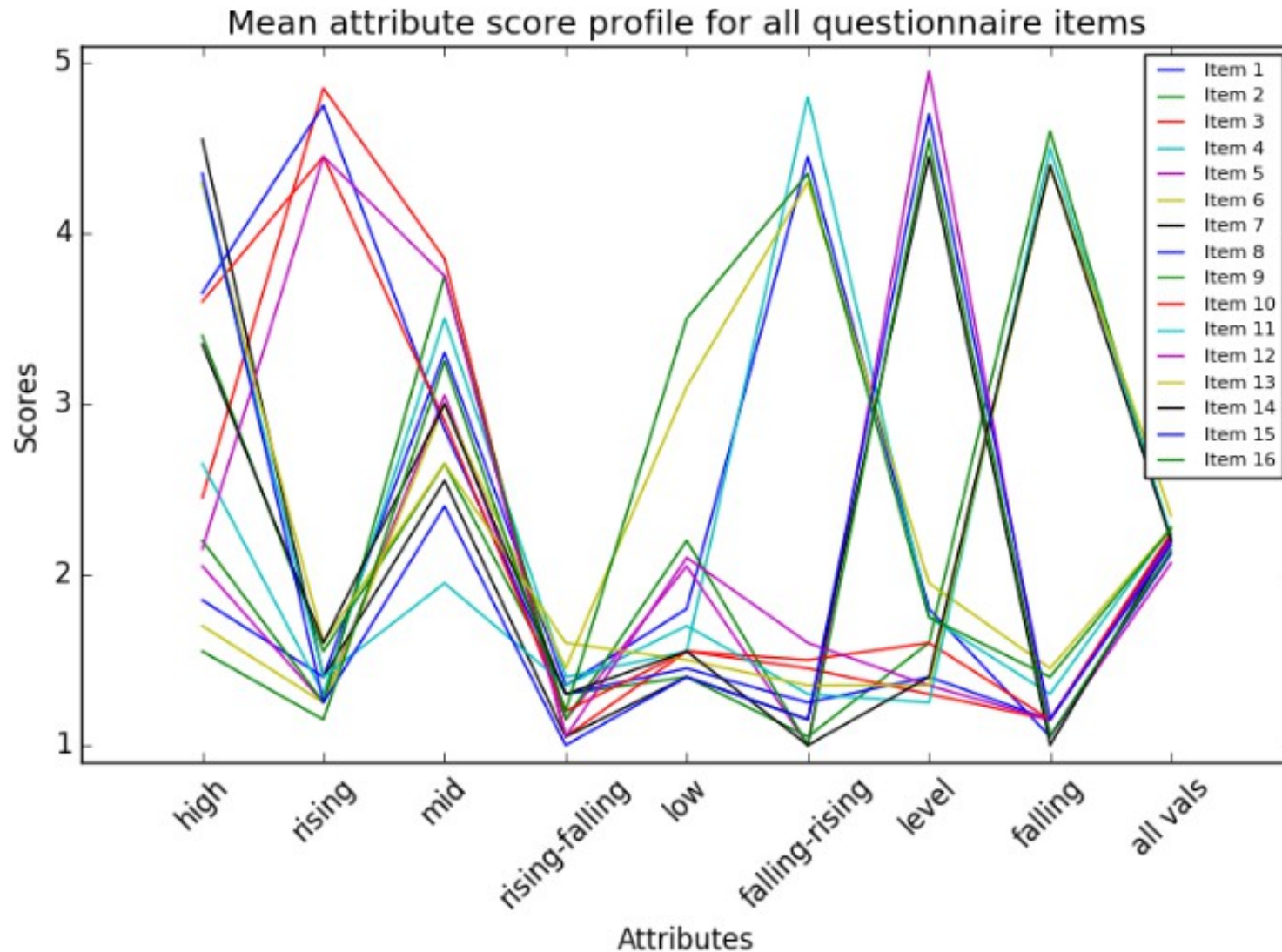
Average value for each attribute for each item



Each line represents the scores for one of the 18 submissions:
high, rising, mid, rising-falling, low, falling-rising, level, falling

(Not very useful unless the responders tend to agree.)

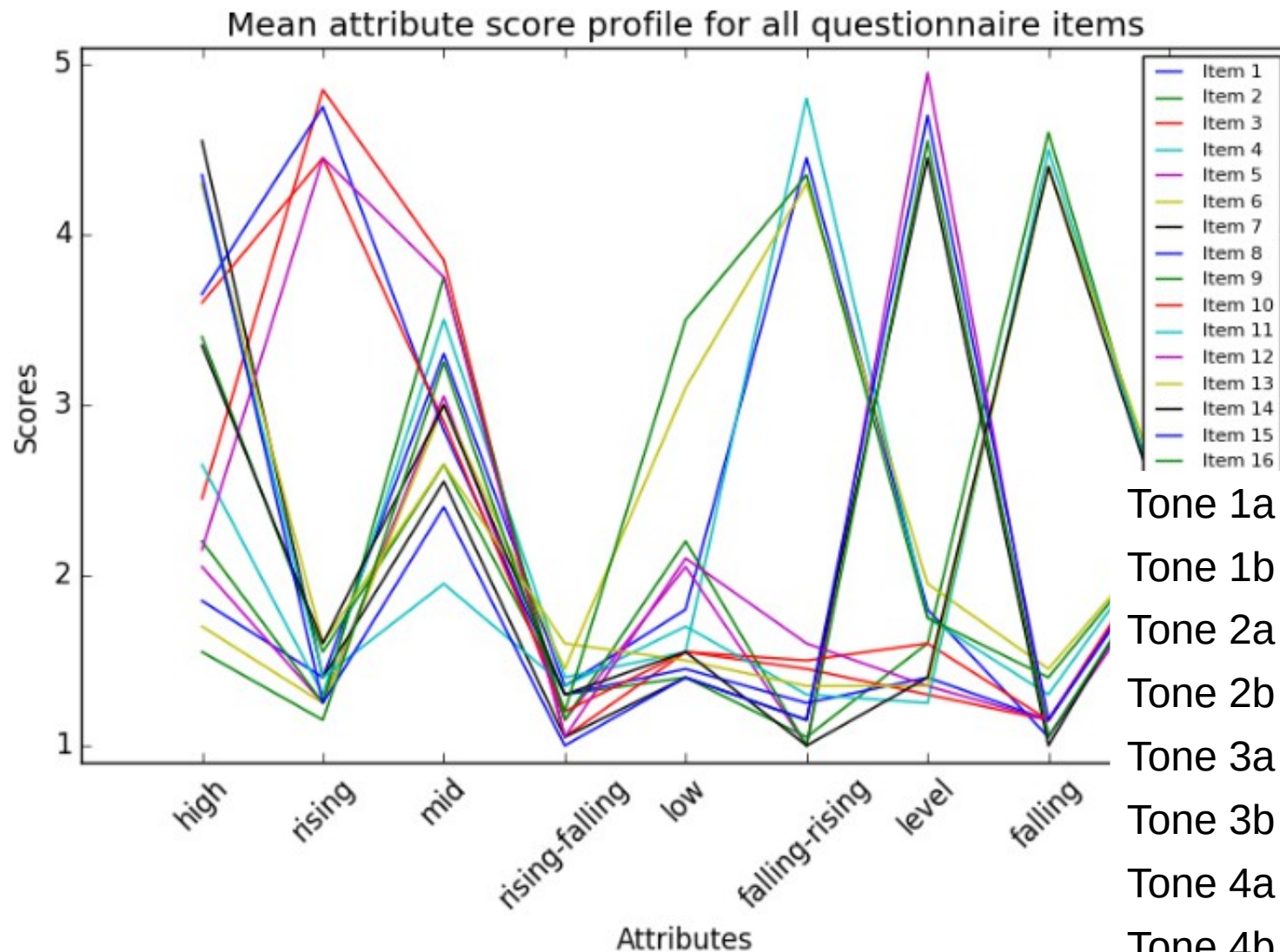
Average value for each attribute for each item



Each line represents the average Likert scores for one of the 16 items.

The similar lines show that some attributes tend to be assigned to a similar extent to all items.

Average value for each attribute for each item



Each line represents the average Likert scores for one of the 16 items.

The similar lines show that some attributes tend to be assigned to a similar extent to all items.

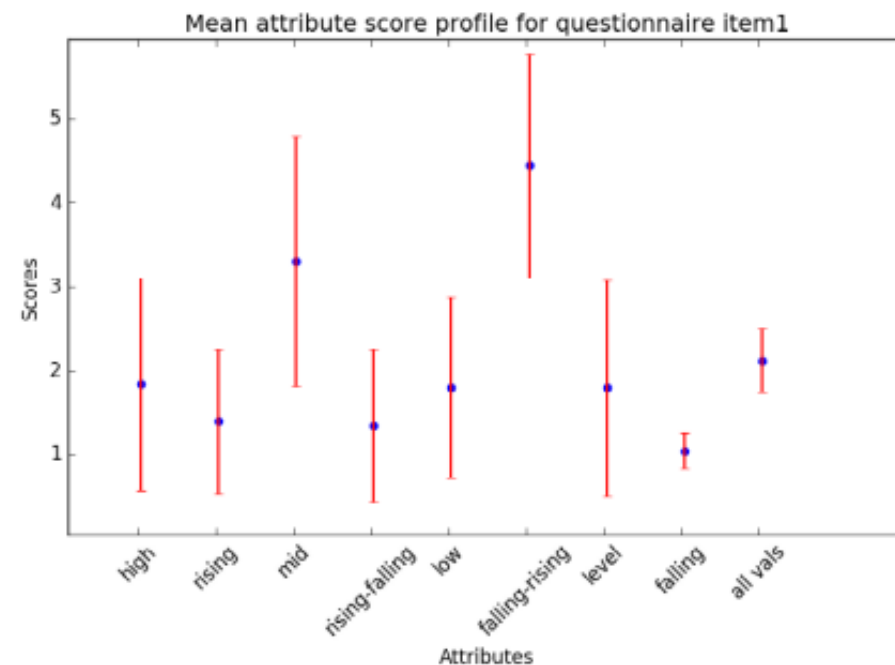
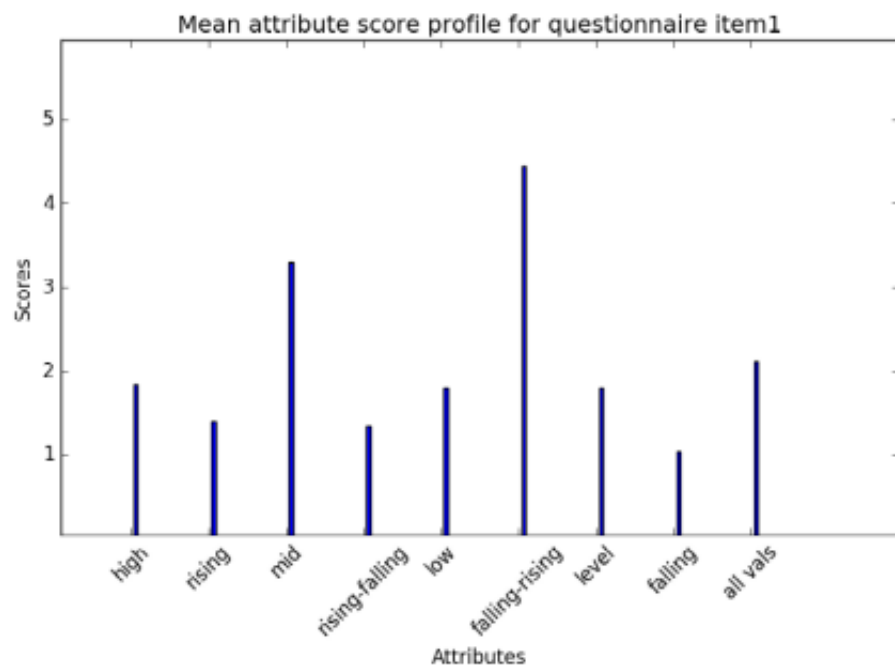
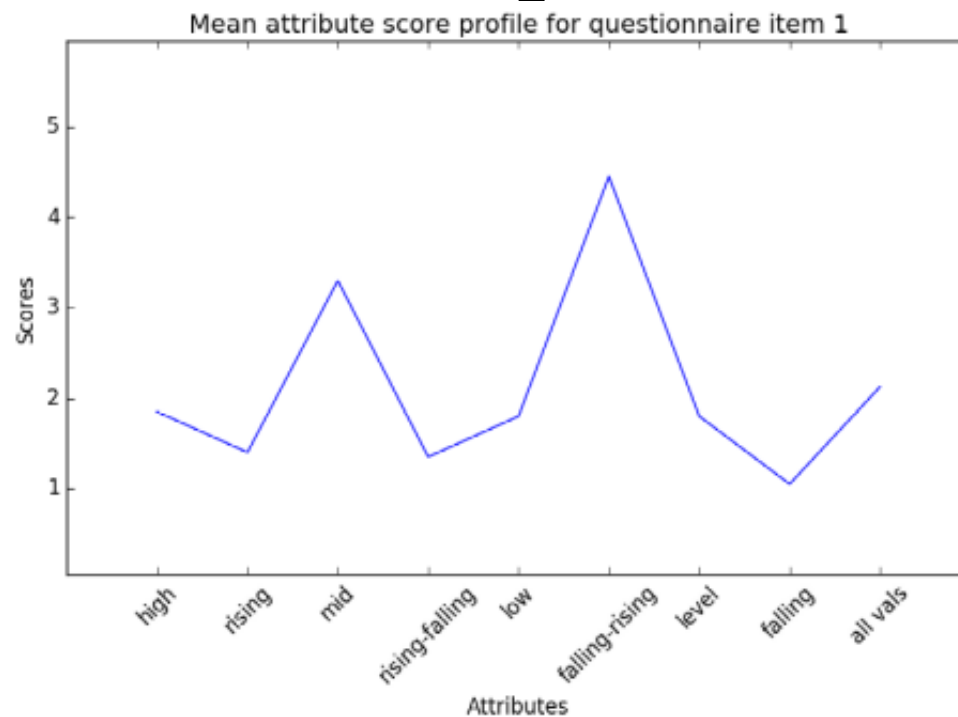
Average value for each attribute for each item

<i>Item</i>	high	rising	mid	rising- falling	low	falling- rising	level	falling
1	1.85	1.40	3.30	1.35	1.80	4.45	1.80	1.05
2	3.40	1.55	2.65	1.30	1.40	1.05	1.60	4.60
3	2.45	4.85	3.85	1.05	1.55	1.45	1.30	1.15
4	2.65	1.30	3.50	1.40	1.55	4.80	1.75	1.30
5	2.05	1.25	3.05	1.15	2.05	1.00	4.95	1.05
6	1.70	1.25	3.00	1.45	3.10	4.30	1.95	1.45
7	4.55	1.40	2.55	1.05	1.40	1.15	4.45	1.00
8	3.65	4.75	2.85	1.30	1.45	1.25	1.40	1.15
9	2.20	1.25	3.75	1.15	1.55	1.00	4.55	1.05
10	3.60	4.45	2.90	Tone 1a	7	15	1.50	1.15
11	4.30	1.40	1.95	Tone 1b	5	9	1.30	4.50
12	2.15	4.45	3.75	Tone 2a	8	10	1.60	1.15
13	4.30	1.60	2.65	Tone 2b	3	12	1.35	4.40
14	3.35	1.60	3.00	Tone 3a	1	4	1.00	4.40
15	4.35	1.25	2.40	Tone 3b	6	16	1.15	4.70
16	1.55	1.15	3.25	Tone 4a	11	13	4.35	1.75
				Tone 4b	2	14		1.40

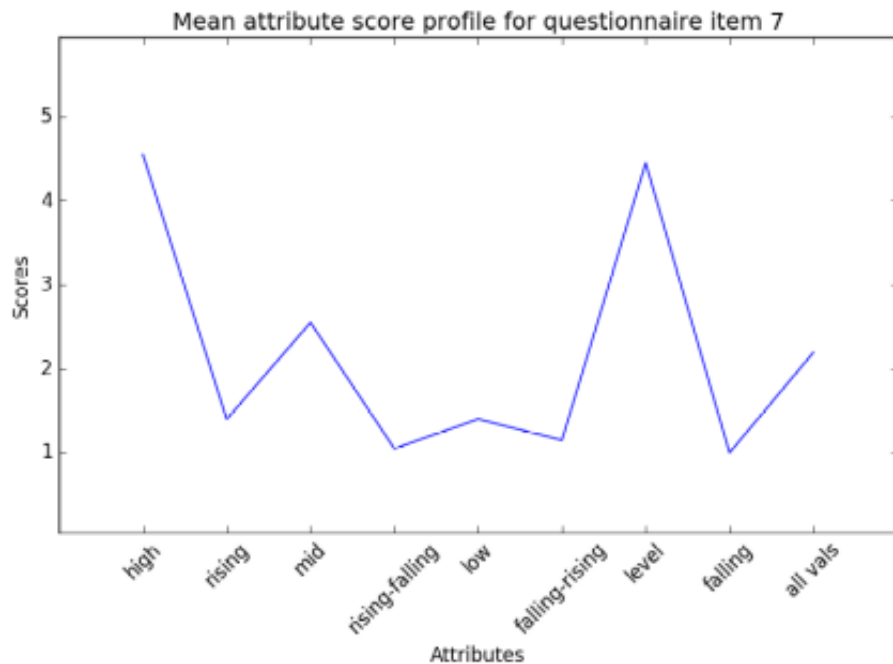
Average value for each attribute for each item

<i>Item</i>	high	rising	mid	rising- falling	low		falling- rising	level	falling
1	1.85	1.40	3.30	Tone 1a	7	15	4.45	1.80	1.05
2	3.40	1.55	2.65	Tone 1b	5	9	1.05	1.60	4.60
3	2.45	4.85	3.85	Tone 2a	8	10	1.45	1.30	1.15
4	2.65	1.30	3.50	Tone 2b	3	12	4.80	1.75	1.30
5	2.05	1.25	3.05	Tone 3a	1	4	1.00	4.95	1.05
6	1.70	1.25	3.00	Tone 3b	6	16	4.30	1.95	1.45
7	4.55	1.40	2.55	Tone 4a	11	13	1.15	4.45	1.00
8	3.65	4.75	2.85	Tone 4b	2	14	1.25	1.40	1.15
9	2.20	1.25	3.75		1.15	2.20	1.00	4.55	1.05
10	3.60	4.45	2.90		1.20	1.55	1.50	1.60	1.15
11	4.30	1.40	1.95		1.35	1.70	1.30	1.25	4.50
12	2.15	4.45	3.75		1.05	2.10	1.60	1.35	1.15
13	4.30	1.60	2.65		1.60	1.50	1.35	1.35	4.40
14	3.35	1.60	3.00		1.30	1.55	1.00	1.40	4.40
15	4.35	1.25	2.40		1.00	1.40	1.15	4.70	1.15
16	1.55	1.15	3.25		1.20	3.50	4.35	1.75	1.40

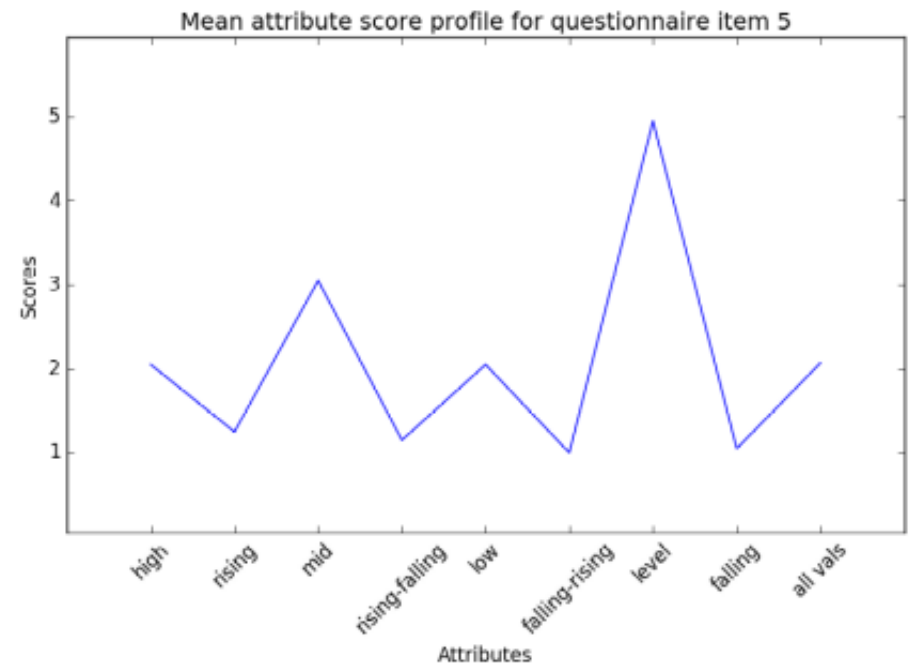
Visualisations of average attribute scores, Tone 3



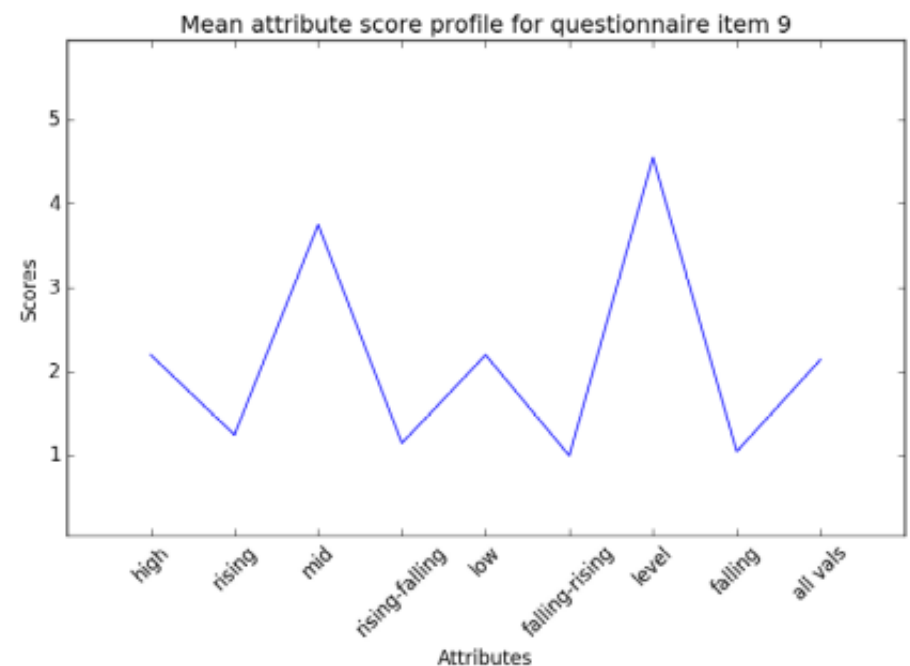
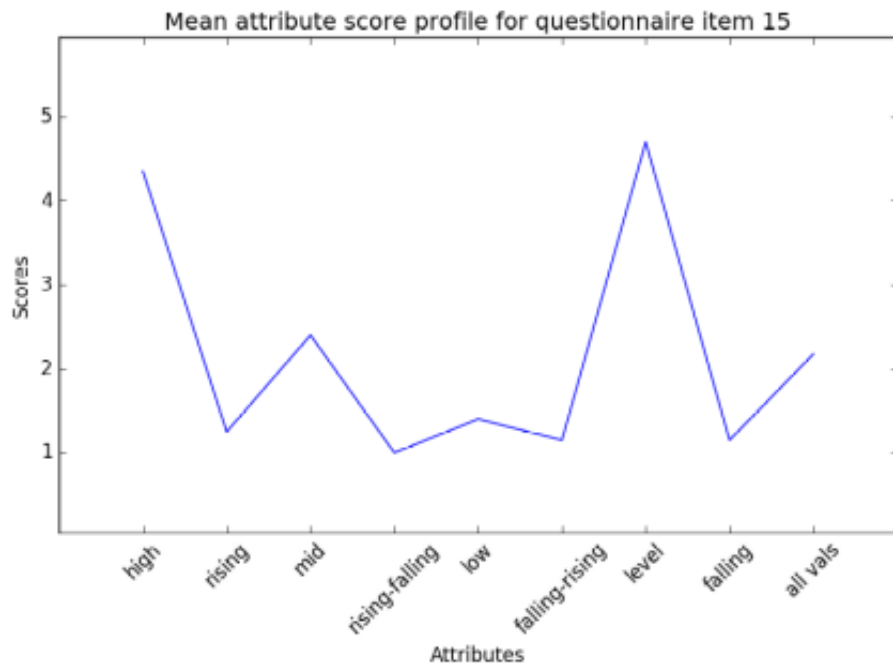
Visualisations of average attribute scores, Tone 1



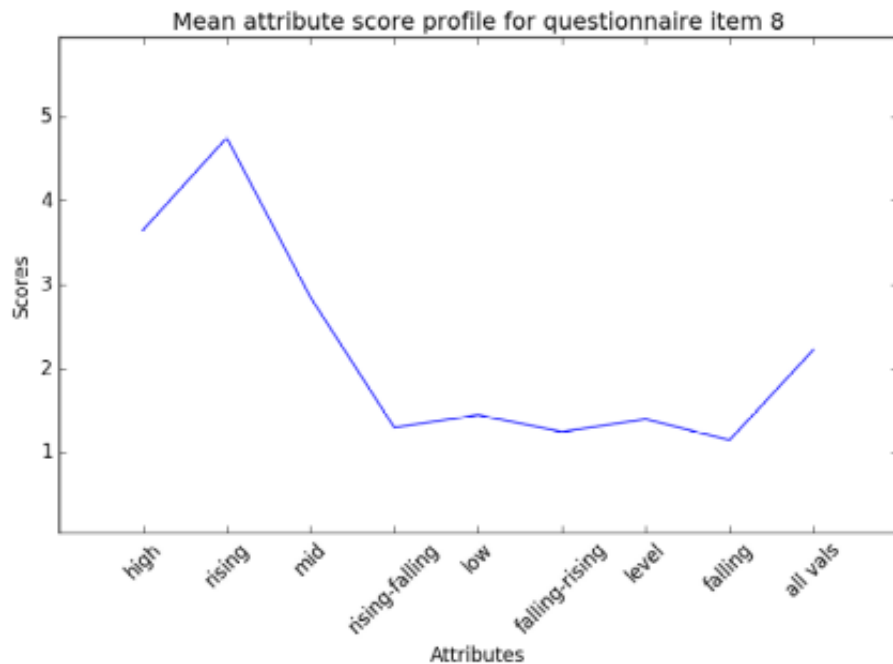
Speaker A



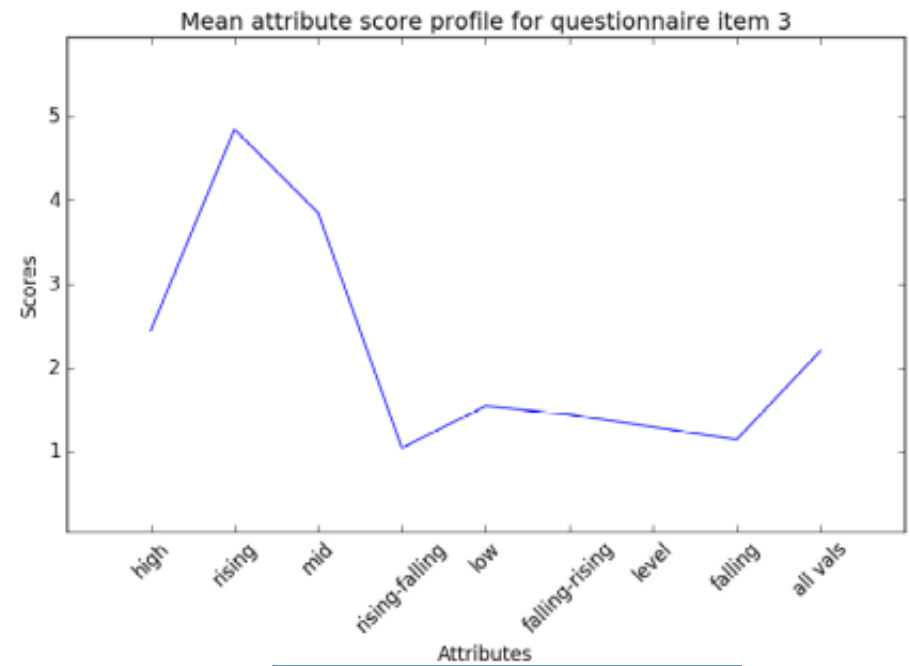
Speaker B



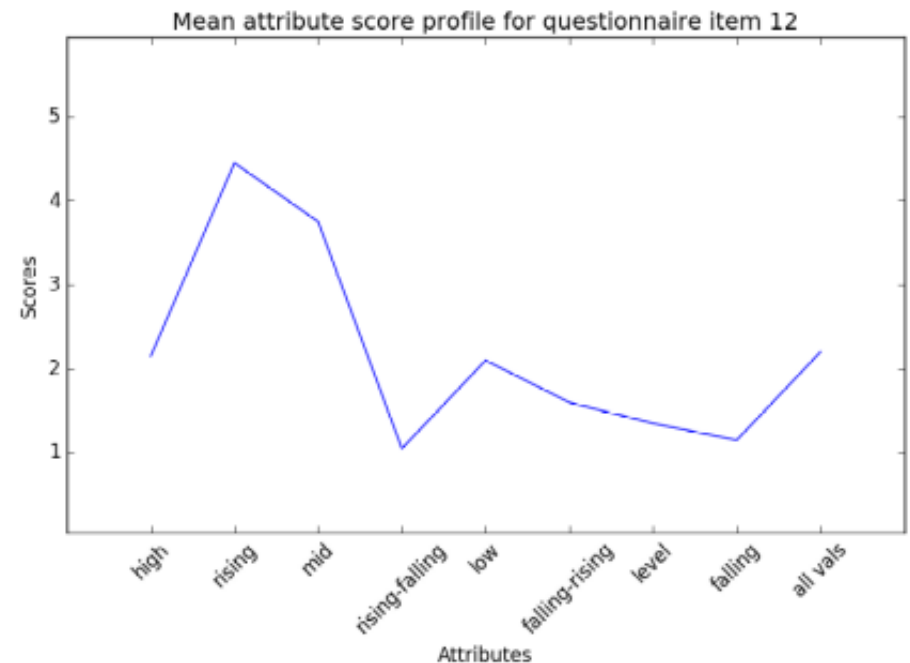
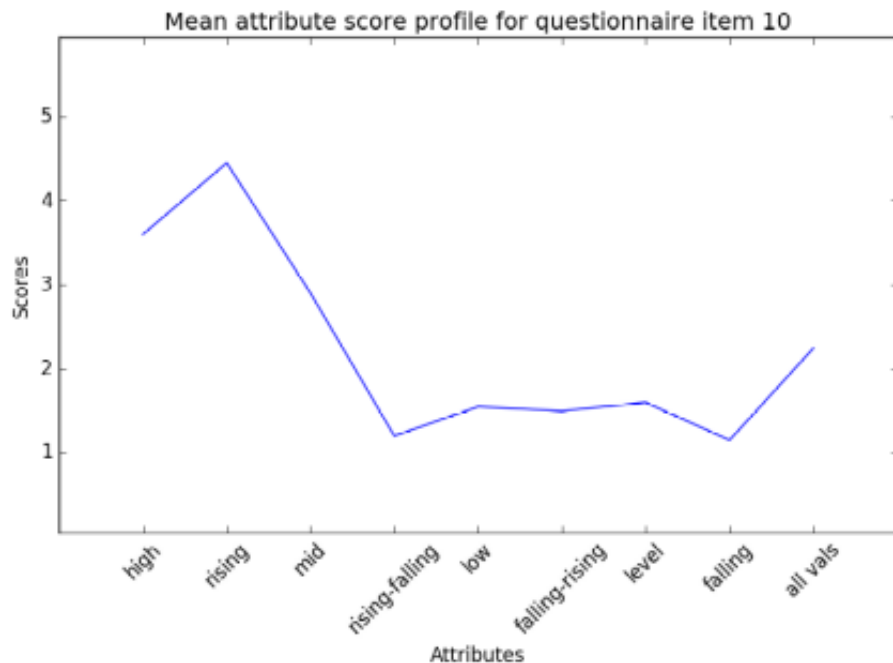
Visualisations of average attribute scores, Tone 2



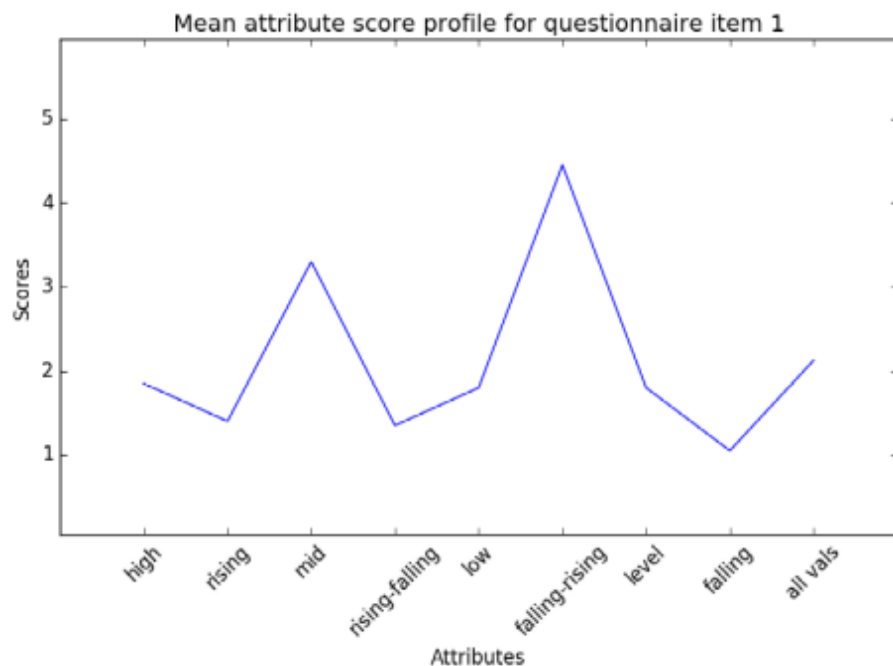
Speaker A



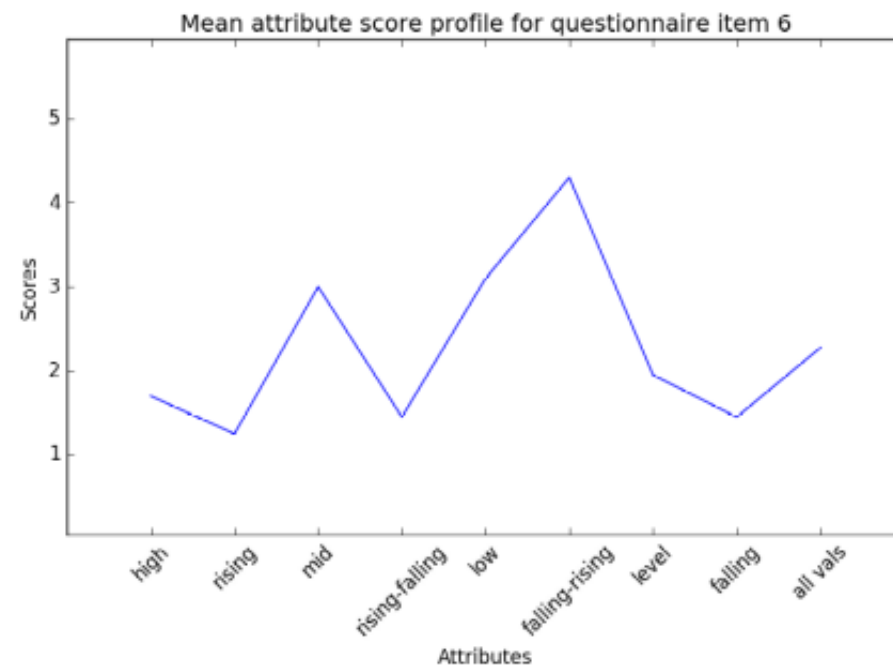
Speaker B



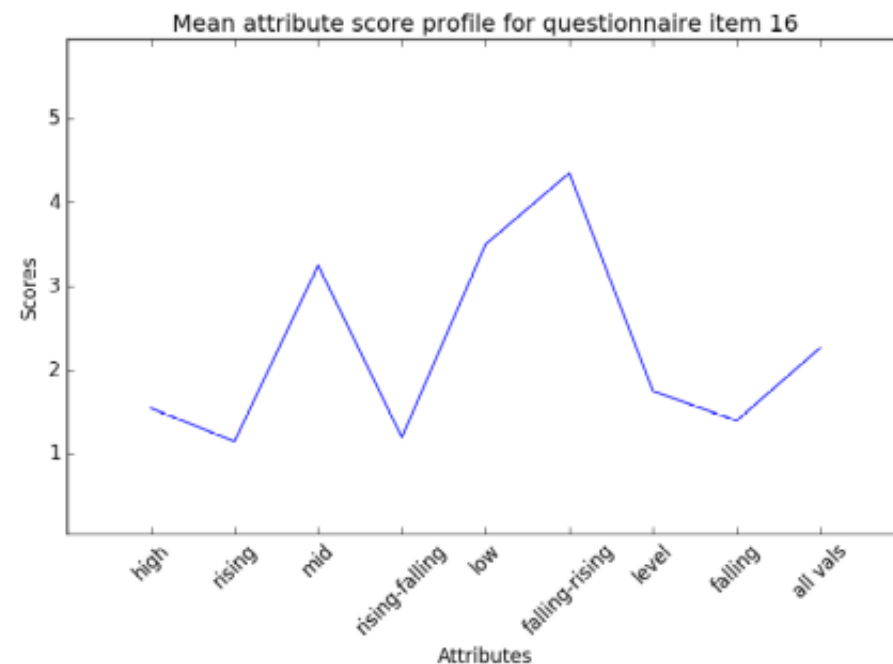
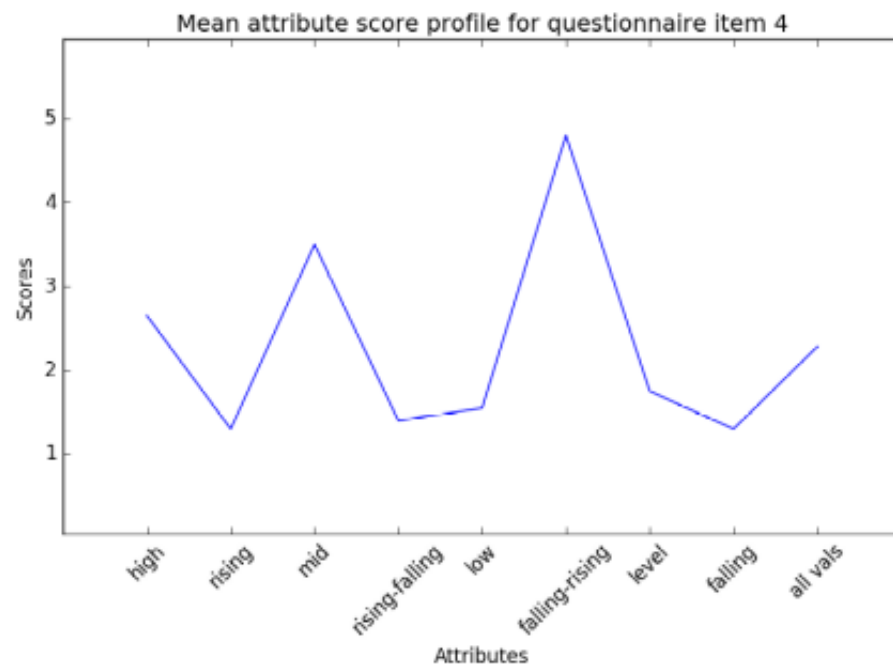
Visualisations of average attribute scores, Tone 3



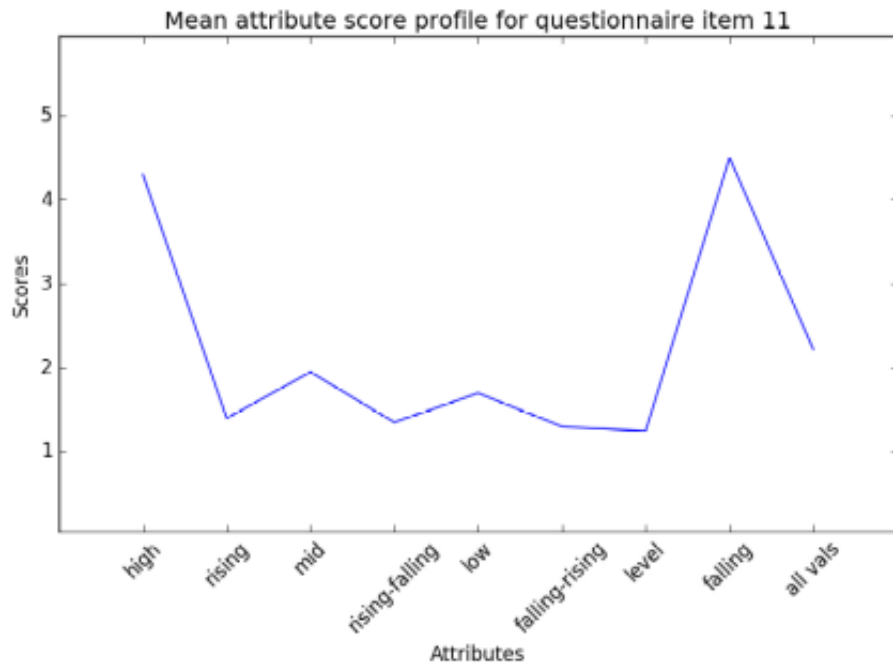
Speaker A



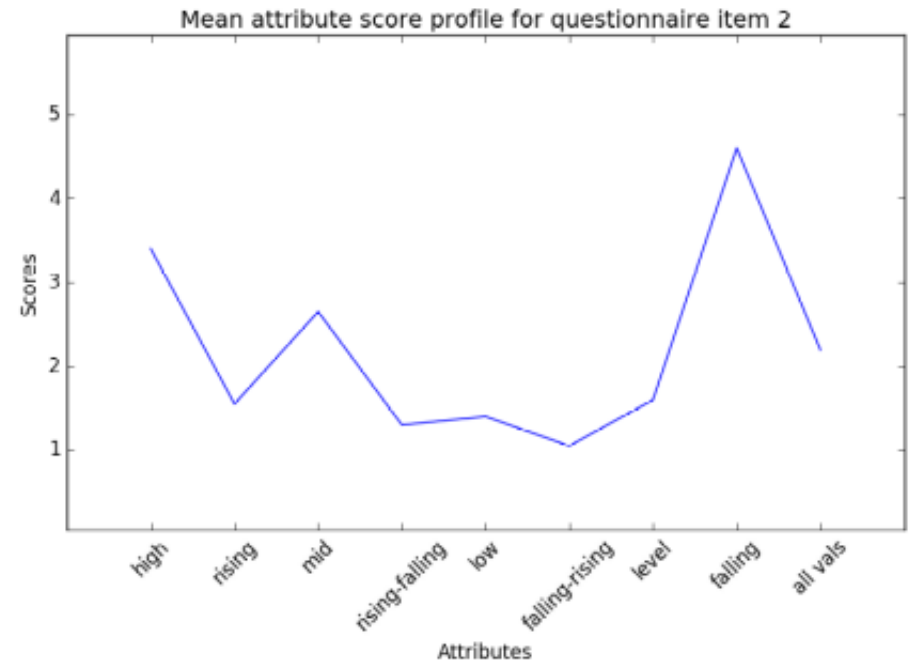
Speaker B



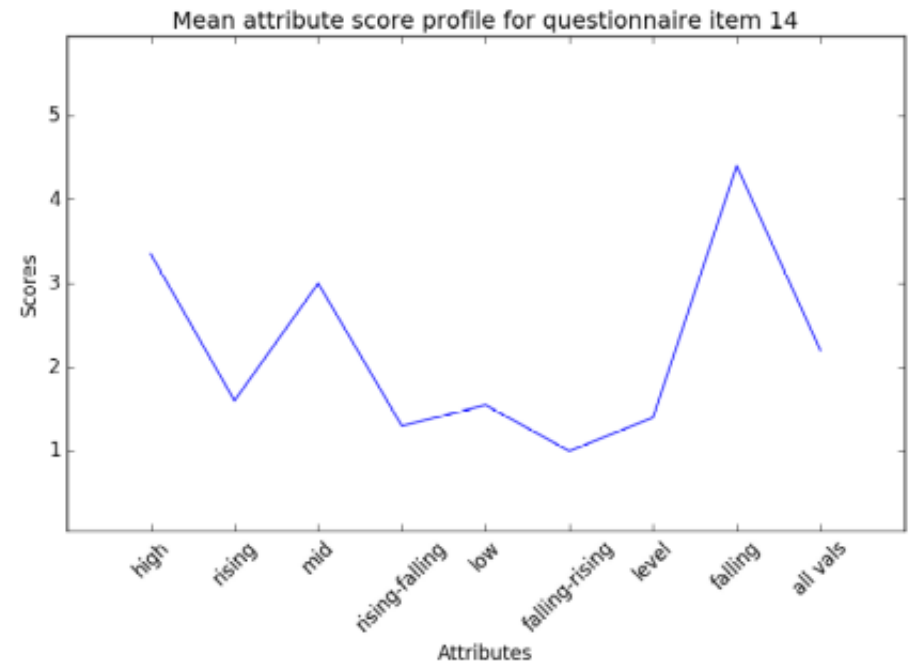
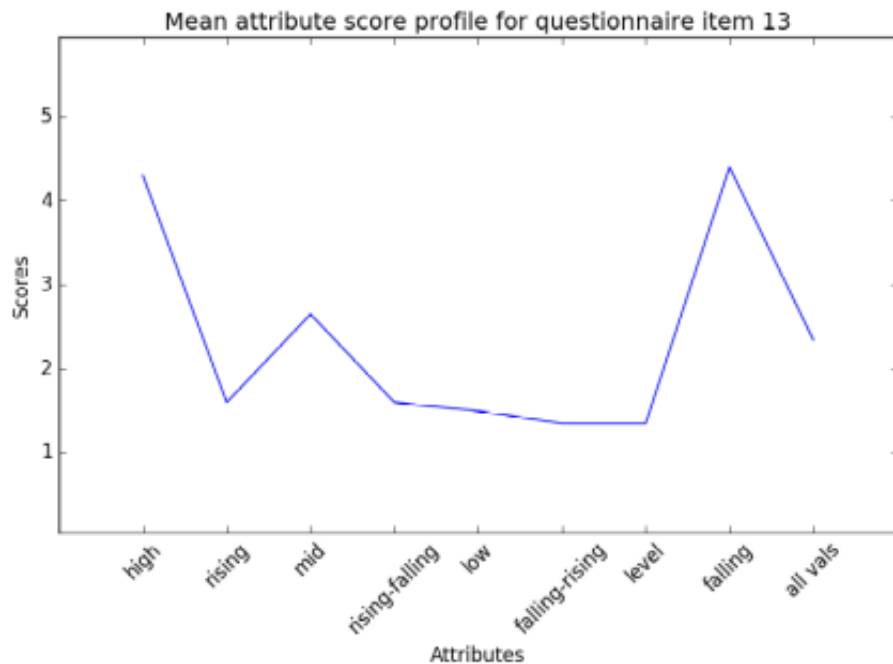
Visualisations of average attribute scores, Tone 4



Speaker A



Speaker B



Free interview survey format: selection of comments

It is normal. The melody is strong and quick. The rhythm is quick. It is relatively slow and mild. The speaker's pitch is high. The rhythm is slow and the speaker is mild. The melody is soft and slow. The speaker's pitch is high.

may be it has a high pitch, and I think she is a young lady. it is quick. Maybe a young lady. it is normal. it is the same as sample one. it is a normal sound. it is slow, I guess the speaker is a middle-aged lady. it is loud. it is strong and powerful.

falling-rising tone, the third of the four tones in modern standard Chinese pronunciation
mid-falling tone, the fourth of the four tones in modern standard Chinese pronunciation
rising tone, the second of the four tones in modern standard Chinese pronunciation
high-falling-rising tone, the third of the four tones in modern standard Chinese pronunciation and high
mid-level tone, the first of the four tones in modern standard Chinese pronunciation
high-falling-rising tone, the third of the four tones in modern standard Chinese pronunciation
high-level tone, the first of the four tones in modern standard Chinese pronunciation
high-rising tone, the second of the four tones in modern standard Chinese pronunciation

Conclusions

- An important type of perception for speech is categorial perception
- There are many ways of documenting perception: survey tests, reaction time tests, transcription, ...
- Perception of tones is basically categorial
 - But with some disagreement on details
- Comments show
 - Phonetic detail judgments
 - Categorical judgments
 - Functional judgments

- Perception, Transcription, Annotation
- Three phases of speech: production, transmission, perception
- Three phases of speech: production, transmission, perception
- Three phases of speech: production, transmission, perception
- Three phases of speech: production, transmission, perception
- Perception – transcription - annotation
- From perception to transcription, from transcription to annotation
- Speech perception
- Speech perception
- Another kind of documentation of perception: opinion survey
- Likert Scale Opinion Survey
- Survey Data Flow
- Questionnaire design and quantitative summary
- Questionnaire evaluation
- Likert scores for all submissions for each attribute
- Likert scores for each attribute/item for all submissions
- Average value for each attribute for each item
- Average value for each attribute for each item
- Average value for each attribute for each item
- Average value for each attribute for each item
- Average value for each attribute for each item
- Visualisations of average attribute scores, Tone 3
- Visualisations of average attribute scores, Tone 1
- Visualisations of average attribute scores, Tone 2
- Visualisations of average attribute scores, Tone 3
- Visualisations of average attribute scores, Tone 4
- Free interview survey format: selection of comments
- Conclusions