## The difference between speech and voice

— Computational description and control of sentiment information embedded in speech —

5 March 2019, at Chulalongkorn univ.

Yoshinori Sagisaka

Waseda University
Applied Math Dept. / GITI /
Language and Speech Science Res. Lab.



1

## Speech and Voice

What is Speech ?

What is the difference between Speech and Voice ?

Voice ⊃ Speech

Speech is Voice containing linguistic information

What is linguistic information ?



2

## Contents of this talk

Description & control for communicative speech

The analysis of N (uhm) to say "Good morning!"
N !! ♫ NN !? N?? N ! NN～?

Sentiment description of speech by color

Can you see and say the color of vowels ?
/a/ /i/ /u/ /e/ /o/



3

## Contents of this talk

Description & control for communicative speech

The analysis of N (uhm) to say "Good morning!"
N !! ♫ NN !? N?? N ! NN～?

Sentiment description of speech by color

Can you see and say the color of vowels ?
/a/ /i/ /u/ /e/ /o/



4

## Studies for synthesis and extraction of communicative information at my lab.

F0 generation for adv+adj speech (Sp. Prosody '04)
- Correlation between comm. F0 and lexicon

Specification of I/O for communicative prosody (ICASSP'05, INTERSPEECH'05, Sp. Prosody '06)
- Input expression by multi-dim. impressions
- Prosody control scheme using lexical attributes

Extraction of speech impression (SNLP'07)

Universality of lexicon-prosody mapping (SNLP'09)

F0 estimation using lexical attributes (SNLP'13)



5

## Contents of this talk

Description & control for communicative speech

- Prosodic differences in communication
- Sentiment description using impressions
- Communicative F0 control using impressions

Sentiment description of speech by color

Can you see and say the color of vowels ?
/a/ /i/ /u/ /e/ /o/



6

## Prosodic differences in communication

Communicative speech has !♫ , ☺ and ●

| | |
|---|---|
| Greetings | Good morning !♫<br>おはようございます !♫<br>สวัสดี !♫ |
| Thanks | Thank you very much ☺<br>どうもありがとうございます ☺<br>ขอบคุณมาก ☺ |
| Apologies | I am very sorry ●<br>どうも すみません ●<br>ฉันเสียใจมาก ● |



7

## Need for Input/Output specification for communicative speech synthesis

To synthesize communicative speech
Thank you very much ☺
どうもありがとうございます ☺
ขอบคุณมาก ☺

We need extra-info ☺ for a synthesis system
- Linguistic content (text) + ?

Beyond text-to-speech synthesis
- Using what input ?
- How to control by what factors ?



8

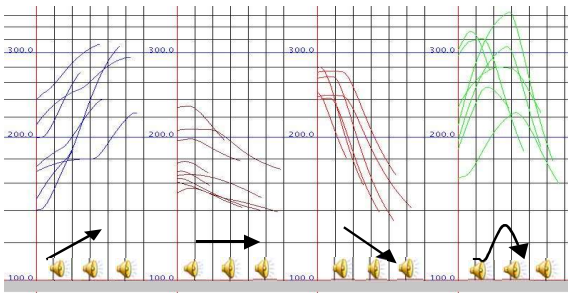
## Information conveyed by prosody

| utterance | information | expressions |
|---|---|---|
| A) N !! ♫ | Praise | (This is nice !) |
| B) NN !? | Doubt | (Really ?) |
| A) N ?? | Incomprehension | (Why not ?) |
| A) N ! | Attention | (Look at this !) |
| B) NN～☺ | Agreement | (I see. ) |



9

## Four typical F0 patterns of “n” observed in real daily conversations






10

## Prototypical F0 patterns by height & dynamics

| Dynamic pattern / height | rise | flat | fall | rise&fall |
|---|---|---|---|---|
| high | Really? ♪<br>Then, what happen?? ♪ ♪<br>Liar!! ♪<br>Really?<br>Is that true? | And? And? ♪ ♪<br>Then what? ♪<br>I agree!!<br>Okay.<br>I don't know…<br>Is that okay? | Reall? I didn't know that!! ♪<br>Of course!! ♪<br>That is nice ♪<br>That is fine.<br>Okay… ↓ | Never mind! ♪<br>That is okay. |
| low | I didn&t know that.<br>Really ↓ | I am not sure…<br>I don't quite agree… | I didn't know that… ↓<br>Yes… but… ↓ ↓ | No, I don't like that ↓ |



11

## Word description for perceptual impression

Speech 12 prototypical single utterances “n”
F0 average height (3) × F0 dynamics (4)

Listeners 5 Adults (2 male, 3 female)

Word description for impression
Possibly followed phrases or words
Imagined speaker’s attitudes



12

## Perceptual impressions

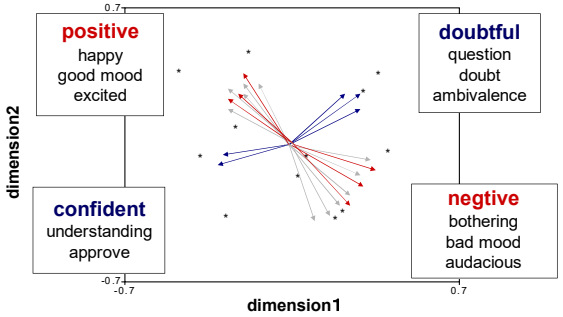
**26 descriptions for perceptual impressions**

1: doubt
2: ambivalence
3: deny
4: question
5: objection
6 approve
7: agreement
8: understanding
9: bright
10: dark
11: cheerful
12:weak
13: interested
14: not interested
15: good mood
16: bad mood
17: light
18: heavy
19: gentle
20: audacious
21: exciting
22: bothering
23: delight
24: anger
25: happy
26: annoying

Waseda University

13

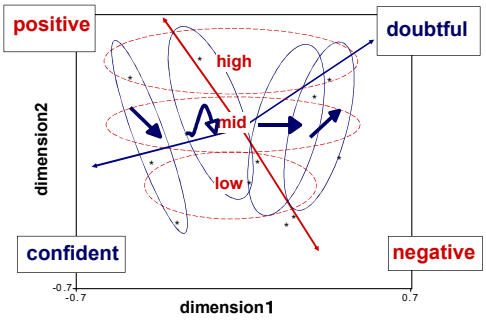
## Word vectors in dim1&dim2-plane




Waseda University

14

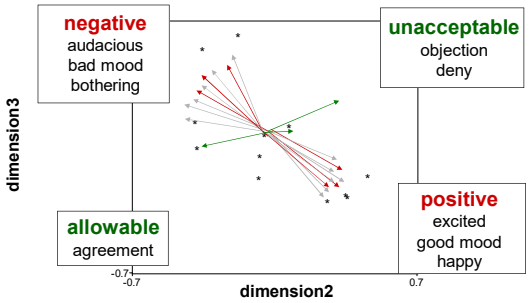
## F0 height and dynamics in dim1&dim2-plane




Waseda University

15

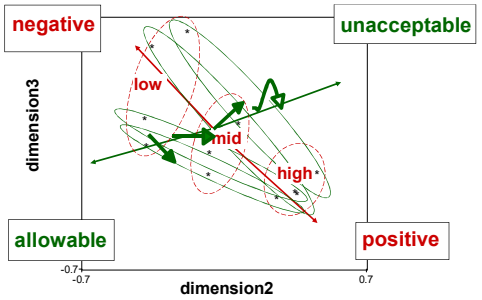
## Word vectors in dim2&dim3-plane




Waseda University

16

## F0 height and dynamics in dim2&dim3-plane

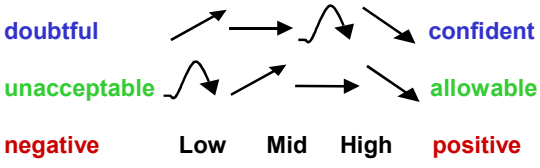



Waseda University

17

## Three dimensional expressions for perceptual impressions

**F0 characteristics corresponds to three dimensional perceptual impressions**
**(Greenberg et al. IEEE ICASSP'05)**




Waseda University

18

## Computation of communicative prosody using input lexical attributes

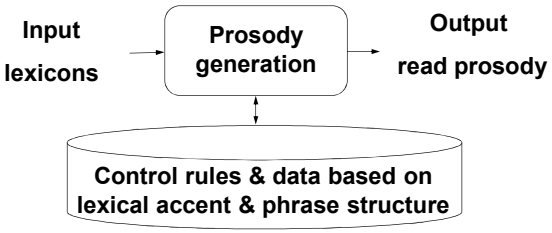
| speaker | utterance | possible expressions |
|---|---|---|
| A) | N !! ♫ | This is nice ! |
| B) | NN !? | Really ? |
| A) | N?? | Why not ? |
| A) | N ! | Look at this ! |
| B) | NN～㋡ | I see. |

Communicative prosody control using lexical attributes
Shao, Greenberg and Sagisaka (SNLP'13)



19

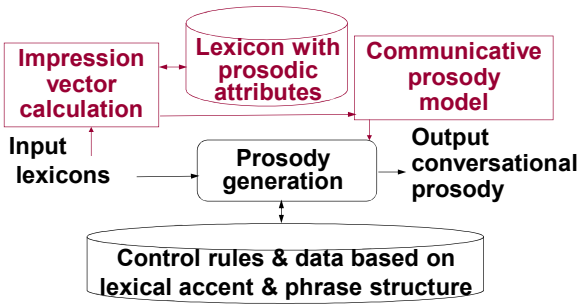
## Prosody generation for test-to-speech






20

## Communicative prosody generation






21

## Example of impression vector computation

| Input | これ | 本当に | きれい | なの |
|---|---|---|---|---|
| | C'est | vraiment | beau | ? |
| | this is | really | beautiful | ? |
| Impression | (degree) | | positive | doubtful |
| value | v1 = +3 | | v2 = +1 | v3 = -4 |

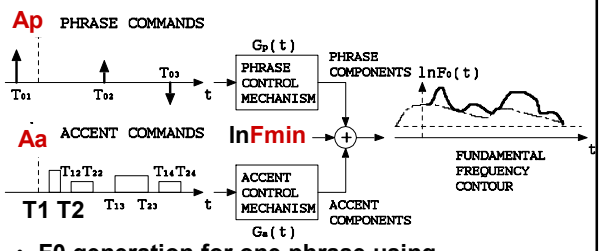
1st dim. doubtful–confident v3 = -4
2nd dim. unacceptable–allowable unassigned 0
3rd dim. negative – positive v1 × v2 = 3

Normalized impression vector = ( -0.8, 0, 0.6 )



22

## F0 generation by command-response model




・ F0 generation for one phrase using 6 parameters (Ap, Aa, lnFmin, T0, T1, T2)



23

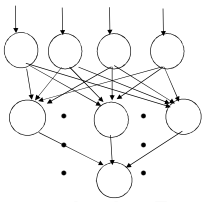
## Communicative F0 control parameter generation from an impression using a NN

Impression vector

Reading F0 generation parameters (Ap Aa Fmin)

Input

Output

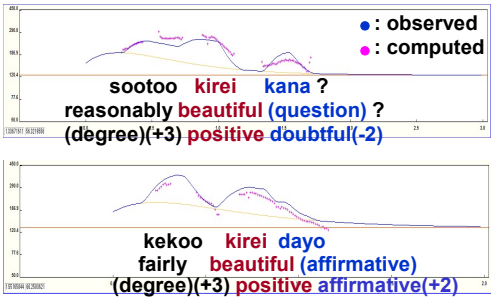


Communicative F0 generation parameters (Ap Aa Fmin)



24

## Computed F0 using lexical attributes




Waseda University

25

## Contents of this talk

Description & control for communicative speech

The analysis of N (uhm) to say "Good morning!"
N !! ♫ NN !? N?? N ! NN～?

**Sentiment description of speech by color**

Can you see and say the color of vowels ?
/a/ /i/ /u/ /e/ /o/

Waseda University

26

## Contents of this talk

Description & control for communicative speech

The analysis of N (uhm) to say "Good morning!"
N !! ♫ NN !? N?? N ! NN～?

**Sentiment description of speech by color**

- Problems in the description of speech & color
- Introduction of sentiment analysis of multi-media
- Sentiment correlations between speech & color
- Communicative speech description using color

Waseda University

27

## Multi-dimensional description for perceptual information

Description using multiply chosen 26 words

**doubtful – confident**
question, doubt, ambivalence, understanding, approve

**unacceptable – allowable**
deny, objection, agreement

**negative – positive**
dark, weakly, not interested, bad mood, heavy, bothering, audacious, anger, annoying, cheerful, delight, gentle, good mood, excited, happy, light, interested, bright

Waseda University

28

## Problems in the description of speech & color

Speech description using impressions
Effective, however,
Fundamentally difficult using discrete symbols

Color has the same description problem using language form
e.g. burgundy red, emerald green, navy blue

→ Limitation using language form

**Mapping between speech and color**
A new possibility for multi-modal info processing

Waseda University

29

## Sentiment analysis studies on correlations between different media

Correlation between images and words
Bouba / kiki effect (W. Köhler 1929)
Sport onomatopoeia (Fujino 2008)

Correlation between sound and color
Non-verbal mapping (Nagata 2003)
Analysis of synesthesia, an ability to hear color

Studies on color association from speech input
Color imaging from vowels (Wrembel 2008)
Mapping from speech to color (Watanabe 2014)

Waseda University

30

## Sentiment analysis studies on correlations between different media

Correlation between images and words
- Bouba / kiki effect (W. Köhler 1929)
- Sport onomatopoeia (Fujino 2008)

Correlation between sound and color
- Non-verbal mapping (Nagata 2003)
- Analysis of synesthesia, an ability to hear color

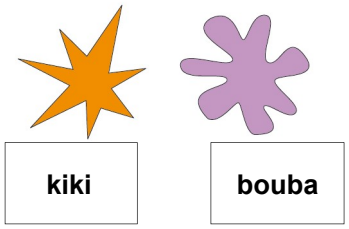
Studies on color association from speech input
- Color imaging from vowels (Wrembel 2008)
- Mapping from speech to color (Watanabe 2014)



31

## Bouba / kiki effect

Which figure corresponds to bouba & kiki ?
Wolfgang Köhler 1929






32

## Sentiment analysis studies on correlations between different media

Correlation between images and words
- Bouba / kiki effect (W. Köhler 1929)
- Sport onomatopoeia (Fujino 2008)

Correlation between sound and color
- Non-verbal mapping (Nagata 2003)
- Analysis of synesthesia, an ability to hear color

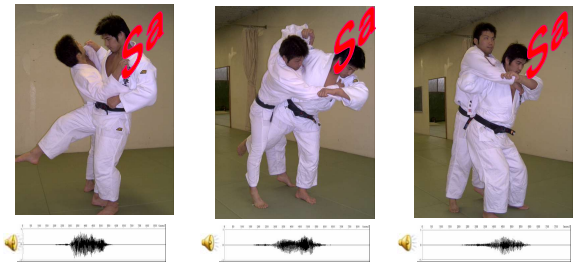
Studies on color association from speech input
- Color imaging from vowels (Wrembel 2008)
- Mapping from speech to color (Watanabe 2014)



33

## Sport onomatopoeia

Different F0 patterns of tsukuri, kuzushi and kake in Judo



Fujino 2008 (Shoogakukan)



34

## Sentiment analysis studies on correlations between different media

Correlation between images and words
- Bouba / kiki effect (W. Köhler 1929)
- Sport onomatopoeia (Fujino 2008)

Correlation between sound and color
- Non-verbal mapping (Nagata 2003)
- Analysis of synesthesia, an ability to hear color

Studies on color association from speech input
- Color imaging from vowels (Wrembel 2008)
- Mapping from speech to color (Watanabe 2014)



35

## Non-verbal mapping between sound & color

Non-verbal mapping sound to color (Nagata 2003)
Compared synesthesia and ordinary people

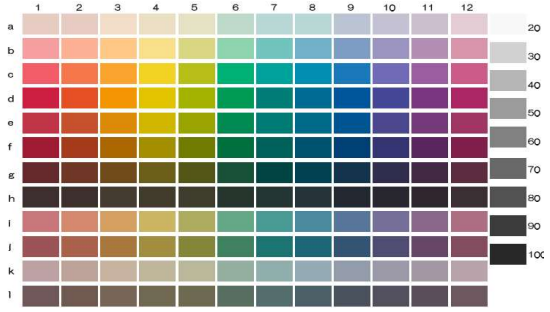
Sound stimuli consisting of
Key : C major to B major, C minor to B minor
Timbre : Pure tone and various types of harmonics
Height : Rising scale of D major

Color selection
153 Hues from
PCCS(Practical Color Co-ordinate System)



36

## Practical Color Co-ordinate System




Waseda University

37

## Sentiment analysis studies on correlations between different media

Correlation between images and words
- Bouba / kiki effect (W. Köhler 1929)
- Sport onomatopoeia (Fujino 2008)

Correlation between sound and color
- Non-verbal mapping (Nagata 2003)
- Analysis of synesthesia, an ability to hear color

Studies on color association from speech input
- **Color imaging from vowels (Wrembel 2008)**
- Mapping from speech to color (Watanabe 2014)

Waseda University

38

## Color imaging from vowels

After listening vowel sounds
Imagined color category was selected

Selected tendency

| Vowel | | Hue |
|---|---|---|
| front /i/ | → | yellow |
| mid /e/ | → | green |
| back /u/ /o/ | → | blown blue black |
| open /a/ | → | red |

(M. Wrembel et al,"Sounds like a rainbow" ISCA WS 2008)

Waseda University

39

## Sentiment analysis studies on correlations between different media

Correlation between images and words
- Bouba / kiki effect (W. Köhler 1929)
- Sport onomatopoeia (Fujino 2008)

Correlation between sound and color
- Non-verbal mapping (Nagata 2003)
- Analysis of synesthesia, an ability to hear color

Studies on color association from speech input
- Color imaging from vowels (Wrembel 2008)
- **Mapping from speech to color (Watanabe 2014)**

Waseda University

40

## Mapping from speech to color

(Watanabe 2014)

After listening vowel sounds
Imagined color category was selected

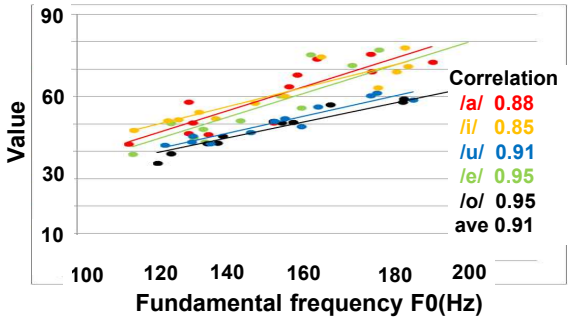
Speech stimuli
Communicative speech consisting of
5 Japanese vowels /a/, /i/, /u/, /e/ and /o/
with 12 F0 patterns (3 heights × 4 dynamics)

Color selection
153 Hues from
PCCS(Practical Color Co-ordinate System)

Waseda University

41

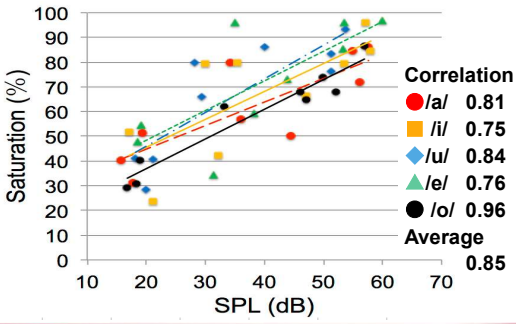
## Correlation between F0 and Value




Waseda University

42

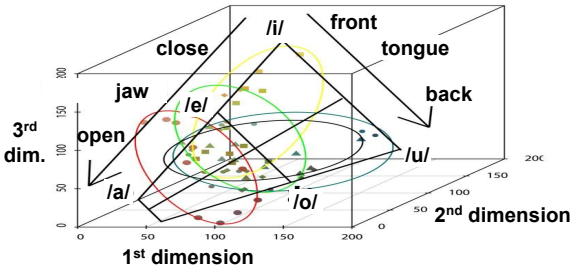
## Correlation between SPL and Saturation




Waseda University

43

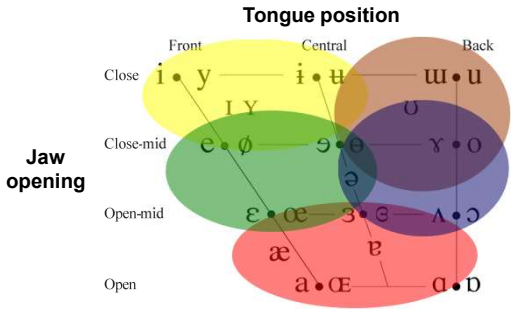
## Correlation between vowels and colors




Waseda University

44

## Correlation between mouth shape and color




Waseda University

45

## Mapping from speech to color

High correlations between
F0s in speech and (Value, Saturation) in Color
0.91 : F0s and Value
0.85 : SPL and Saturation

Same Vowel - Hue correlation in Wrembel 2008

| Vowel | | Hue |
|---|---|---|
| front /i/ | → | yellow |
| mid /e/ | → | green |
| back /u/ /o/ | → | blown blue black |
| open /a/ | → | red |

Waseda University

46

## Summary

**Description & control for communicative prosody**

- Impression description of speech
- 3 dimensional impressions for F0 control
- Communicative F0 computation scheme

**Sentiment description of speech by color**

- Introduction of multi-media info interactions
- Highly correlations are found between F0 and Value, SPL and Saturation & Vowel and Hue

Waseda University

47

## What I would like to discuss with you

**Is speech merely superficial form of language?**

- Linguistic treatment of communicative info?
- What is linguistics?
- No need of the redefinition of linguistic info?

**Generalized linguistic information?**

- Freedom from written language
- Information descriptions using X-media
- Pan-linguistic information

Waseda University

48

## Acknowledgements

Thanks to all the current and previous lab members

Special thanks to collaborators

Yoko Greenberg
Yoshotaka Fujino
Lu Shao
Kanako Watanabe