Multiple Acoustic Cues for Stop Recognition

Weonhee Yun
University of Edinburgh

Outline

- Typology of Korean stops
- Acoustic cues for stops
- Distribution of acoustic parameters
- Statistical analysis (MANOVA and Post-hoc test)
  - Speech with contextual variability
- System overview
- Results and discussion

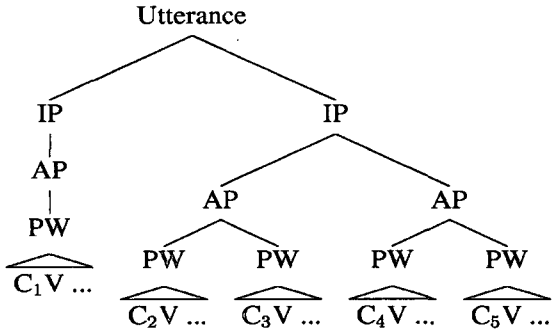
Typology of Korean Stops

Type	Phoneme	Example
Lax	p, t, k	불, 달, 기
Tense	p', t', k'	뿔, 딸, 끼
Aspirated	p^h, t^h, k^h	풀, 탈, 키

Acoustic Cues for Stops

- Voice Onset Time (VOT)
  - » Tense < Lax < Aspirated
- Closure duration
  - » Lax < Aspirated < Tense
- Fundamental frequency (F0) of a vowel after a stop
  - » Lax < Tense <Aspirated

Prosodic Position and Domain-Initial Strengthening



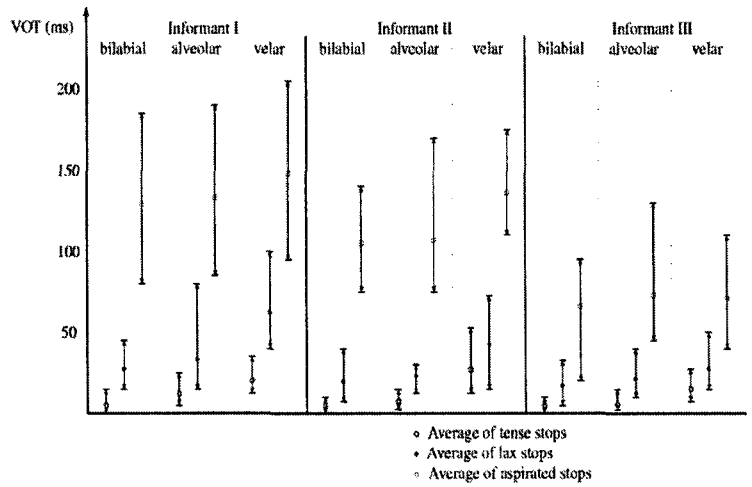
VOT

Aspirated stops:	$C_1 > C_2 > C_4 > C_3$ (or C_5)
Lax stops:	$C_1 > C_2 > C_4 > C_3$ (or C_5)
Tense stops:	No systematic difference

Closure duration

Aspirated stops:	$C_2 > C_4 > C_3$ (or C_5)
Lax stops:	$C_2 > C_4 > C_3$ (or C_5)
Tense stops:	$C_2 > C_4 > C_3$ (or C_5)

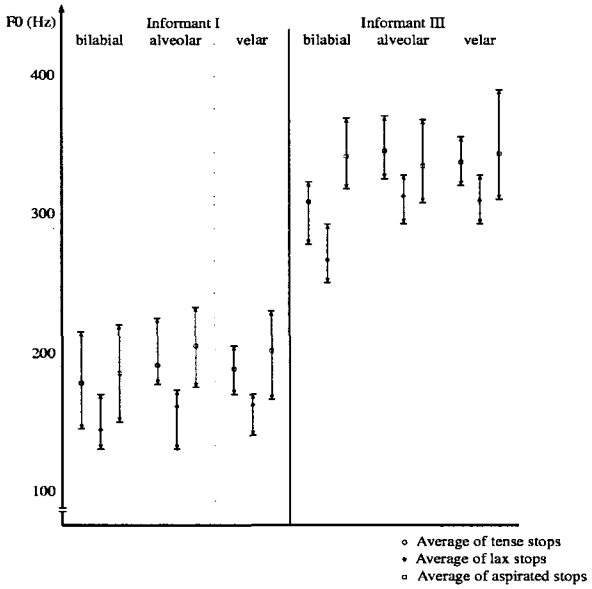
VOT (Han & Weitzman 70)



Closure duration (Silva 92)

Prosody	Phrase Edge		Word Edge		Word Internal	
Preceding	Vowel	Nasal	Vowel	Nasal	Vowel	Nasal
p	66	54	50	32	48	25
p^h	85	73	77	53	84	57
p'	104	98	81	60	123	83
Mean	85	75	69	48	85	56

F0 after stops



Statistical Tests on Speech with Contextual Variability

- Effect of
  - Phonation type
  - Place of articulation
  - Prosodic position
  - Preceding phone context

Phonation Type and Place of Articulation

		Labial		Alveolar		Velar	
		Significance Level		Significance Level		Significance Level	
		$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
CLS	ASP-LAX	√		√	√	√	√
	ASP-TNS	√	√	√	√	√	√
	LAX-TNS	√	√	√	√	√	√
VOT	ASP-LAX			√	√		
	ASP-TNS	√	√	√	√	√	√
	LAX-TNS	√	√	√	√	√	√
F0	ASP-LAX	√	√	√	√	√	√
	ASP-TNS	√	√			√	√
	LAX-TNS	√	√	√	√	√	√

Prosodic Word Initial

		Labial		Alveolar		Velar	
		Significance Level		Significance Level		Significance Level	
		$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
CLS	ASP-LAX					√	√
	ASP-TNS	√	√				
	LAX-TNS	√	√			√	√
VOT	ASP-LAX			√			
	ASP-TNS	√	√	√	√	√	√
	LAX-TNS	√	√	√	√	√	√
F0	ASP-LAX	√	√	√	√	√	√
	ASP-TNS	√	√			√	√
	LAX-TNS	√	√	√		√	√

Prosodic Word Medial

		Labial		Alveolar		Velar	
		Significance Level		Significance Level		Significance Level	
		$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.01$
CLS	ASP-LAX	√	√	√	√	√	√
	ASP-TNS	√	√	√	√	√	√
	LAX-TNS	√	√	√	√	√	√
VOT	ASP-LAX	√	√	√	√	√	√
	ASP-TNS	√	√	√	√	√	√
	LAX-TNS			√	√		
F0	ASP-LAX			√	√		
	ASP-TNS					√	√
	LAX-TNS			√			

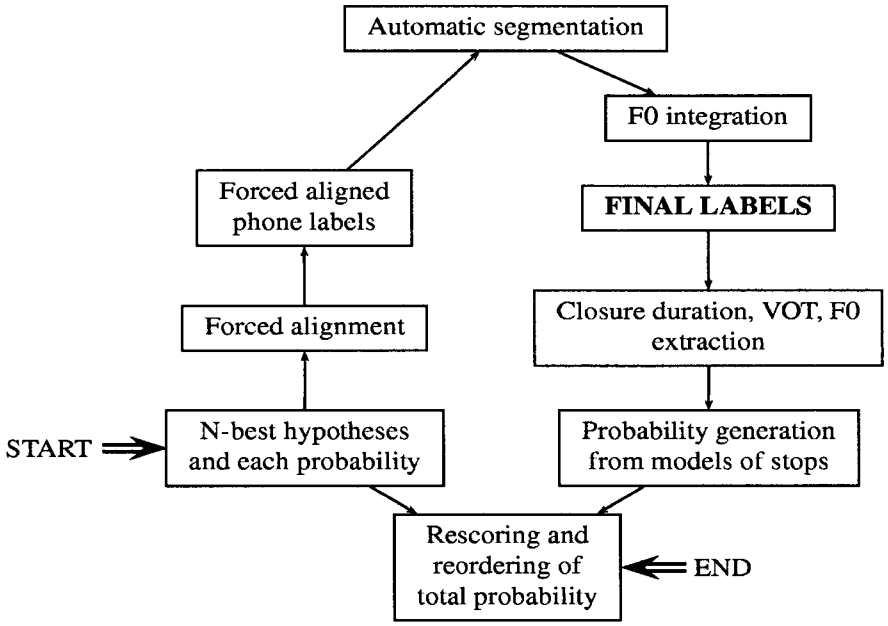
Preceding Phone Context

		LAX-TNS			TNS-ASP			ASP-LAX		
		Stop	Vowel	Son	Stop	Vowel	Son	Stop	Vowel	Son
LAB	CLS	na.	√√	√√	na.	√√	√√	na.	√√	
	VOT	na.	√√	√√	na.	√√	√√	na.		
	F0	na.	√√		na.		√	na.	√√	√√
ALV	CLS	na.	√√	na.				na.	√√	√
	VOT	na.	√√	na.	√√	√√	√√	na.	√	
	F0	na.	√√	na.				na.	√√	√√
VEL	CLS	na.	√√	√√	na.	√√		na.	√√	
	VOT	na.	√√	√√	na.	√√	√	na.		
	F0	na.			na.	√√	√√	na.	√√	√√

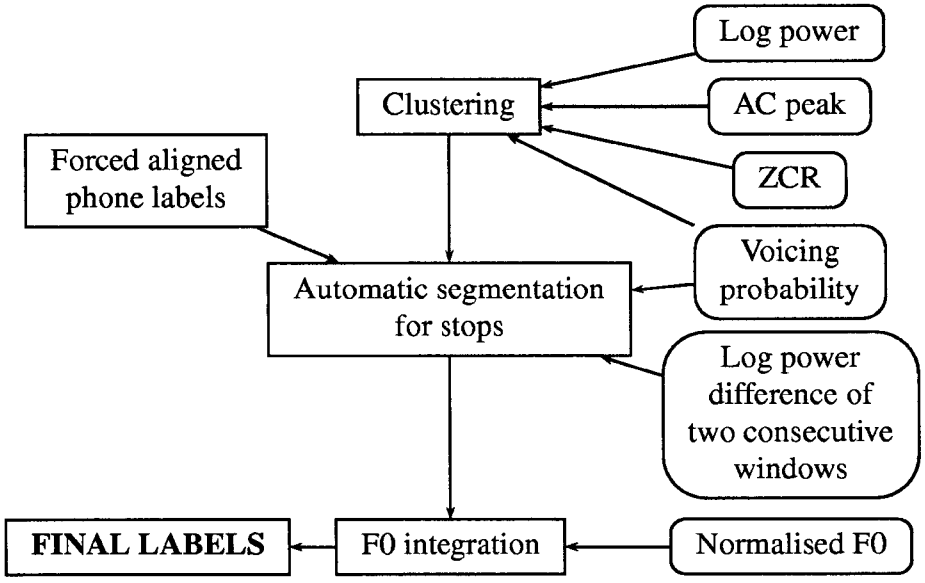
Summary

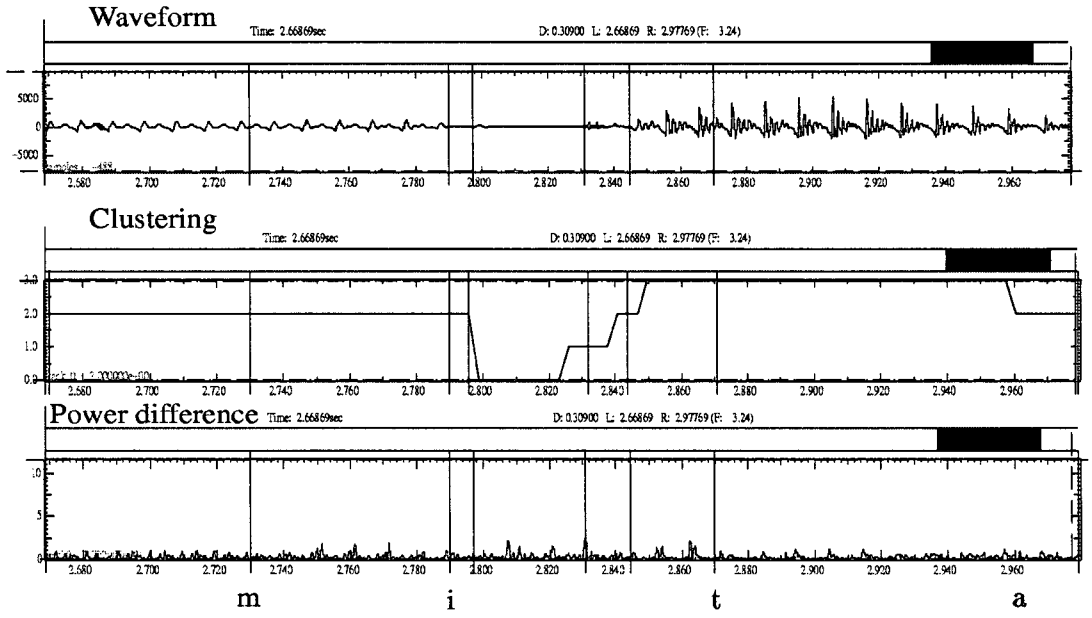
- Acoustic characteristics are maintained
- No domain-initial strengthening effect
- To differentiate phonation types, acoustic cues should be used together

System Overview



Segmentation of Closure and VOT





Modelling Acoustic Parameters

- Parametric family of continuous pdfs
  - Univariate pdfs
    - Normal pdf
    - Gamma pdf
  - Multivariate normal pdf
- Root Mean Squared Error

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(M_i - x_i)^2}{n}}$$

where M_i is a value predicted from the model, which is either a gamma or a normal pdf.

Post-Processing of Stop Probability

- Total probability of each hypothesis

$$P_{total} = P_{N-best} + \alpha P_{stop}$$

- Stop probability

$$P_{stop} = \frac{\sum_{i=1}^{num} P_{stop_i}}{num}$$

Univariate pdf $\quad P_{stop_i} = P_{closure} + P_{VOT} + P_{F0}$

Multivariate pdf $\quad P_{stop_i} = P_{(closure,VOT,F0)}$

Baseline Models

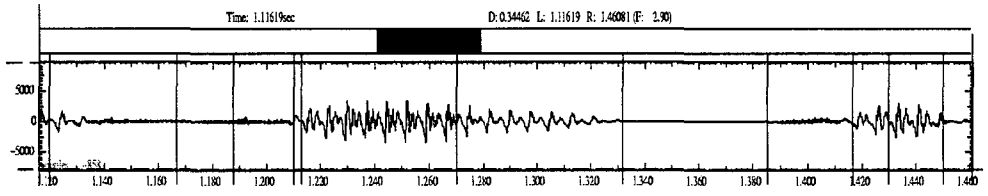
	Choi *et al* (1995)	Yun *et al* (1997)	Jang (2000b)	Our model
No sp	65.3		78.47	79.74
With sp	69.4	68.00		89.01
C.D. no sp	89.5		87.93	88.11
C.D. with sp	90.7	91.11		**95.56**
Demi. no sp			91.03	
Demi. with sp				91.98

N-best Recognition Result

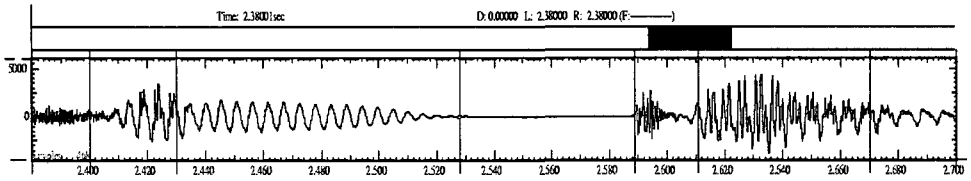
n-best	Word accuracy (%)	Sentence correctness (%)
1st	95.56	75.92
2nd	97.49	86.58
3rd	98.11	89.85
4th	98.38	91.43
5th	98.54	92.21
⋮	⋮	⋮
9th	98.97	94.40
10th	99.01	94.54

Results

	Gamma pdf		Multivariate normal pdf	
Post-processing results	Word acc. (%)	Sentence (%)	Word acc. (%)	Sentence (%)
All 1st best	95.59	75.99	95.56	75.95
Num.	2965		2965	
Reordered dataset only	Word acc. (%)	Sentence (%)	Word acc. (%)	Sentence (%)
Before	65.85	0.00	83.75	33.33
After	82.93	40.00	82.50	33.33
Num.	5		9	



	Gamma pdf			Multivariate normal pdf
	Closure	VOT	F0	All in one
1st ([k])	0.0547	0.2243	1.057e-5	0.0127
2nd ([p^h])	0.1012	0.2599	8.984e-5	0.0840



h a n cls k ə

1st hypothesis

h a ŋ cls k' ə

2nd hypothesis

	Gamma pdf			Multivariate normal pdf
	Closure	VOT	F0	All in one
1st ([k])	0.0478	0.1278	9.311e-7	0.0020
2nd ([k'])	0.0500	0.2123	3.869e-5	0.1883

Correct vs. Substituted Words

- Correct Substitution
  - 변동 변경
  - 배달 대답
  - 동경 공장
  - 도와 보여
  - 불어 팔월

Analysis

- 36 sentences have at least one substituted word involving stops
  - 14 phonation type confusion
  - 16 place of articulation confusion
  - 6 phonation and place confusions
- Not many stops in the test data
- Pseudomorpheme dictionary increases the number of minimal pairs (POSTECH dictionary: 23.58 %)

Summary

- Acoustic characteristics of stops in speech with contextual variability
- Possibility of stop recognition by post-processing technique
- Further work
  - Speech database
  - Modification of decoder
  - Automatic segmentation of acoustic parameters