

# Automatic Assessment of Prosody

Kyuchul Yoon

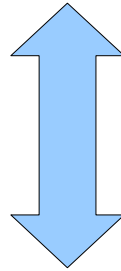
School of English Language & Literature  
Yeungnam University, South Korea

# GOAL

Evaluate  
prosodic(suprasegmental)  
proficiency  
of non-native learners of English

# Prosodic Evaluation

Native Speaker's Utterance



Learner's Utterance

# Comparison between...

(1) F0 contours

(2) Intensity contour

(3) Segmental durations

of the two utterances of the same sentence

# But first...

Make matching segmental durations

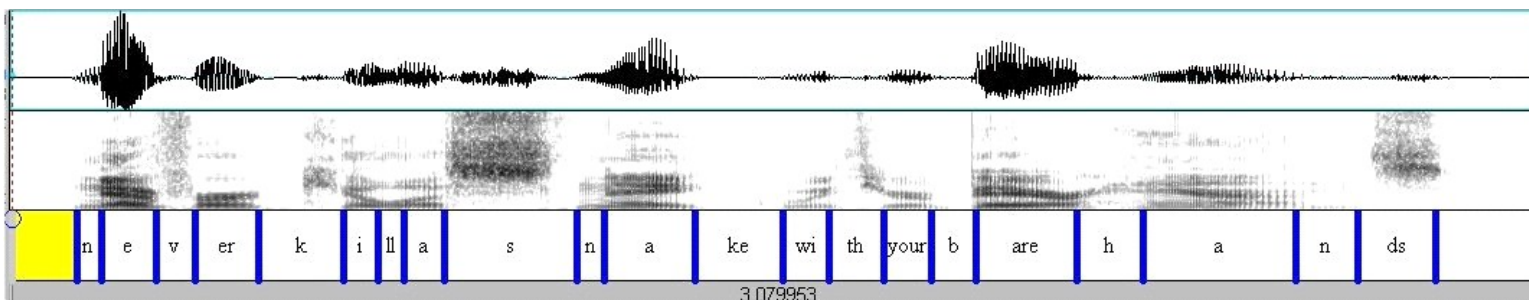
**identical**

so that

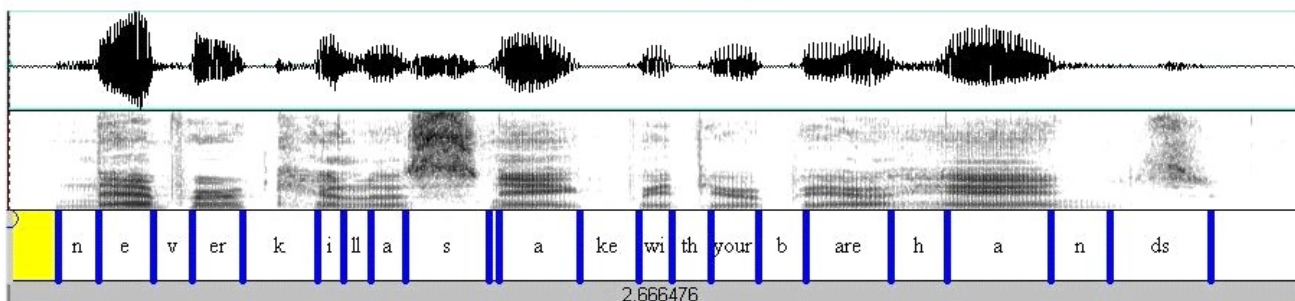
F0 contours & intensity contours can be compared directly.

# Segmental Duration Syncing

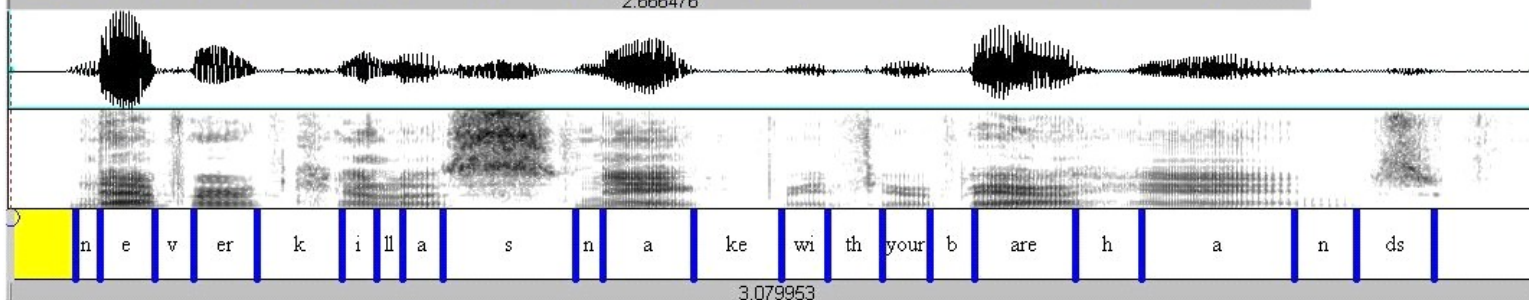
**native**



**learner  
before**



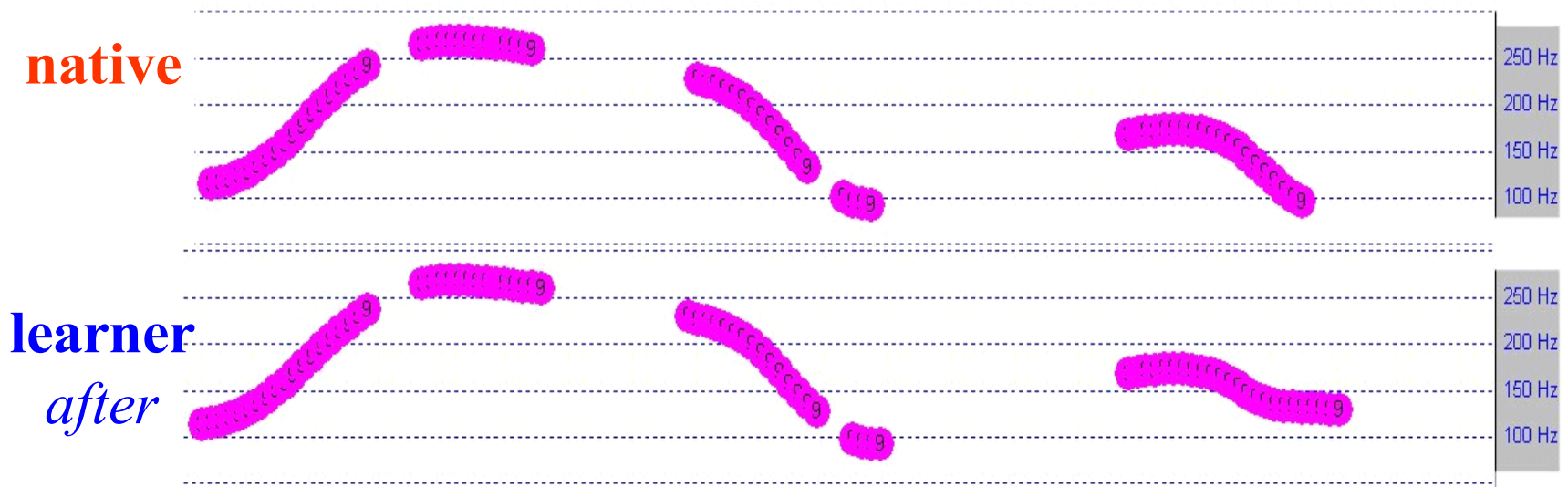
**learner  
after**



Use of PSOLA algorithm (Moulines & Charpentier, 1990)

# Comparison: F0 Contour

F0 : point-to-point comparison btw/ native and learner  
after normalization

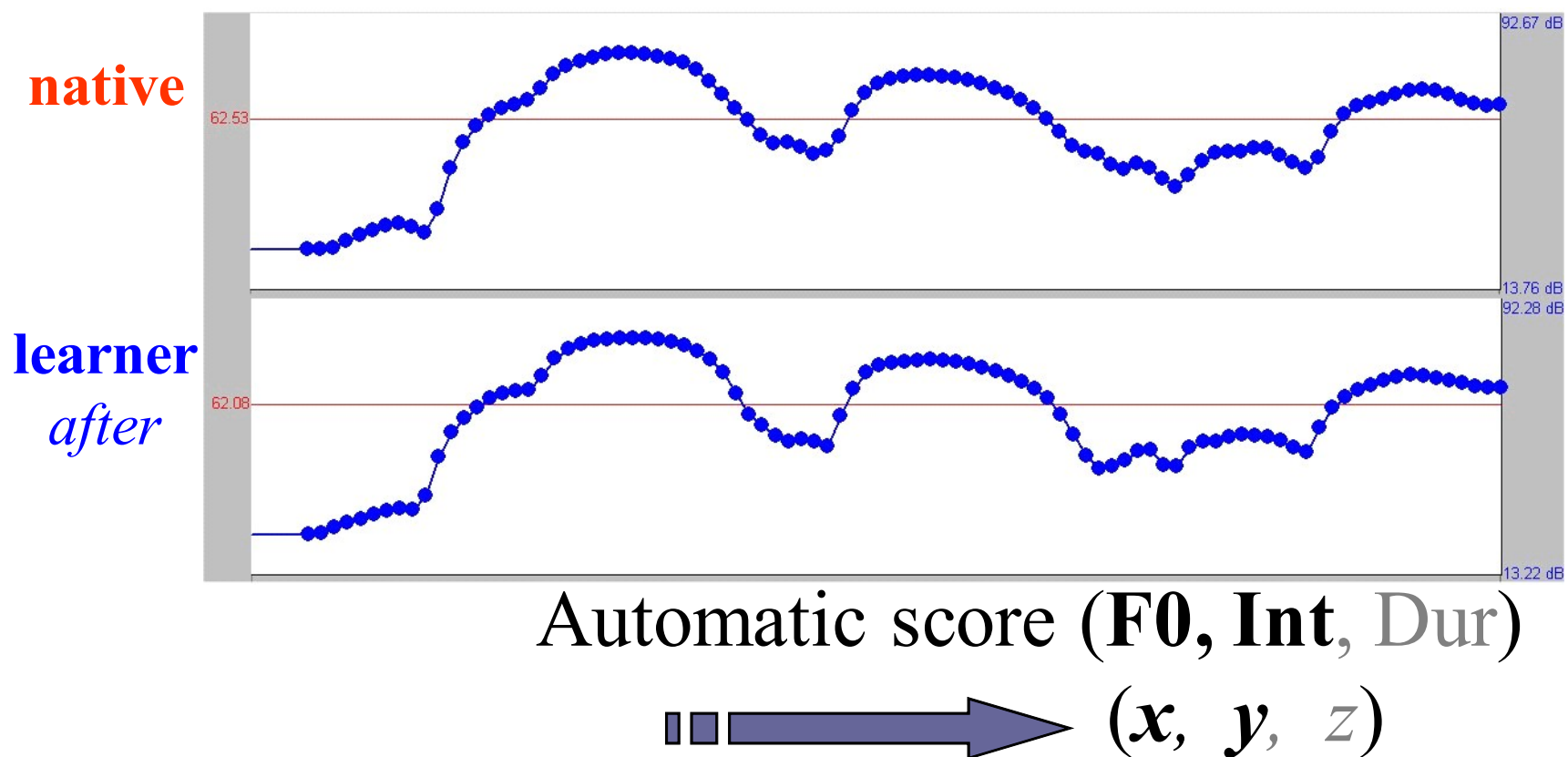


Automatic score (**F0**, Int, Dur)

■ ■ → (**x**, **y**, **z**)

# Comparison: Intensity Contour

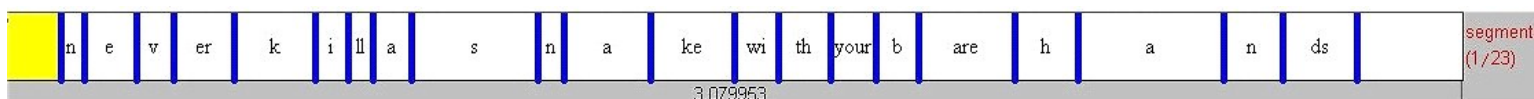
Intensity : point-to-point comparison btw/ native and learner after normalization



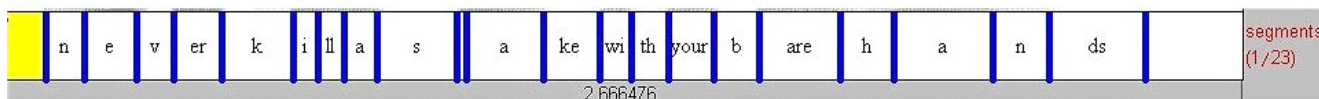
# Comparison: Segmental Durations

Duration : segment-to-segment comparison btw/ native and learner

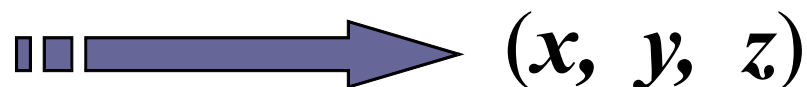
**native**



**learner**  
*before*



Automatic score (**F0, Int, Dur**)



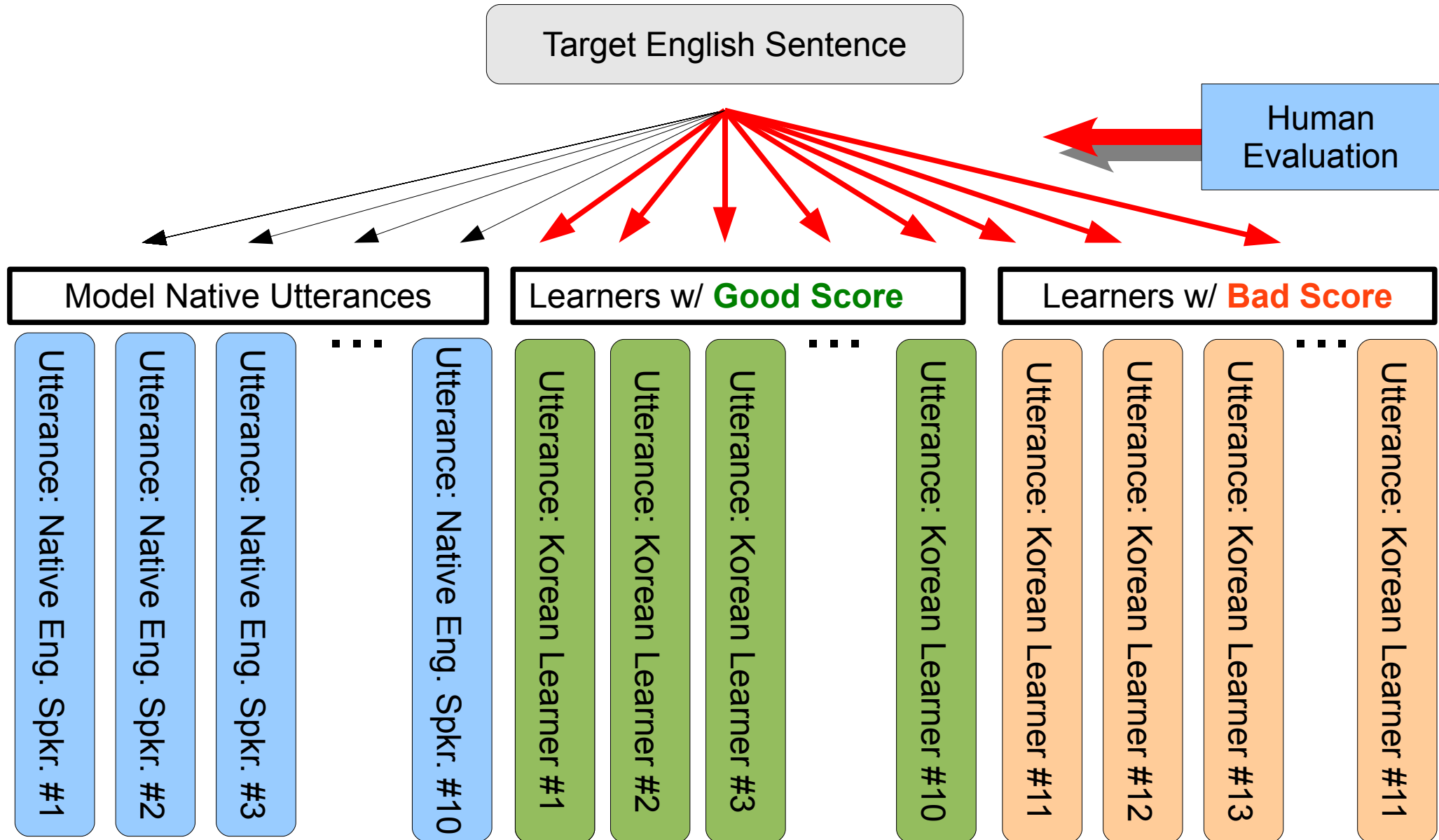
# Metric Used

Euclidean distance metric for evaluation measure

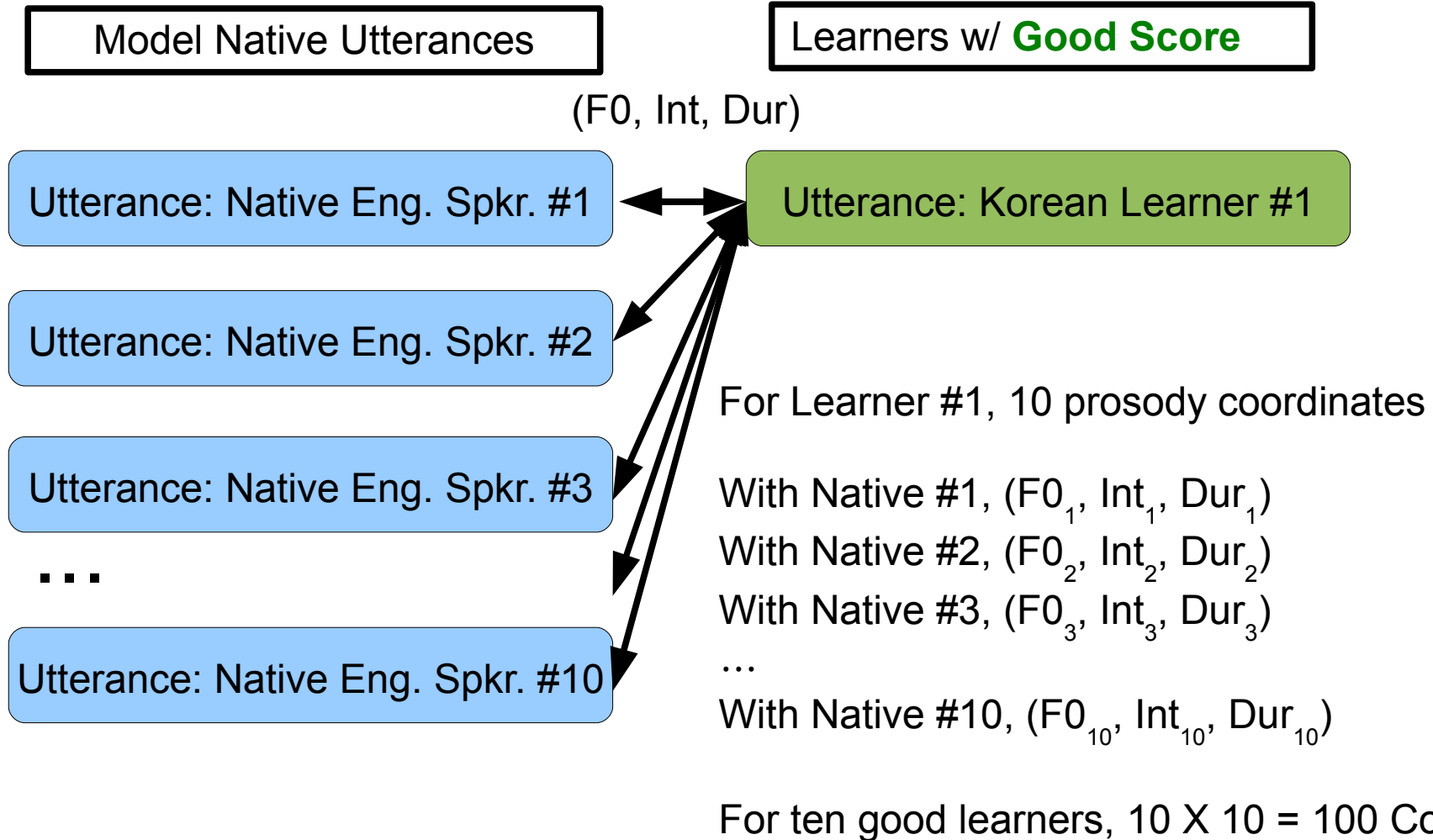
$P = (p_1, p_2, p_3, \dots, p_n)$  and  $Q = (q_1, q_2, q_3, \dots, q_n)$  in Euclidean  $n$ -dimensional space

$$\sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}.$$

# How to Compare?

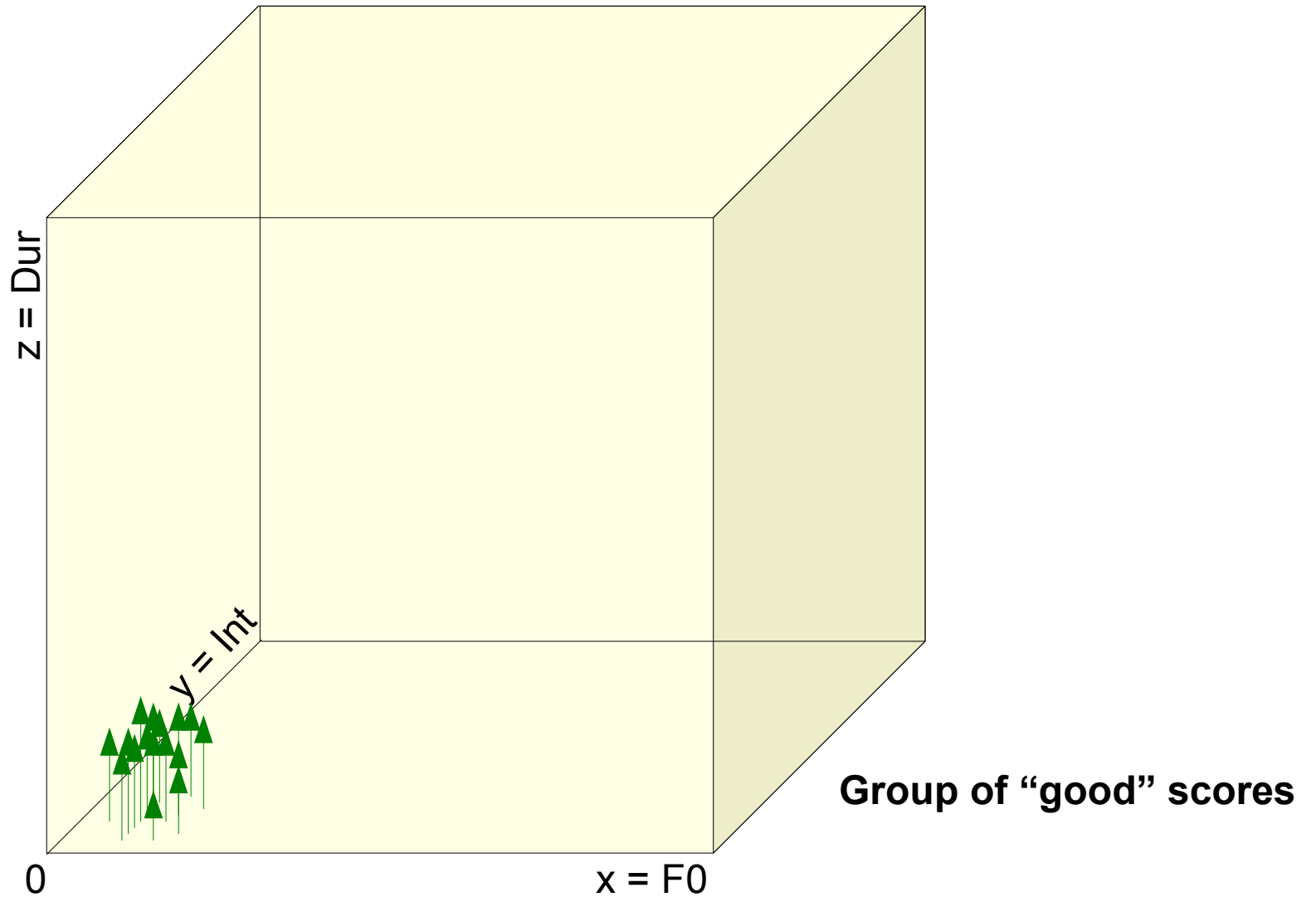


# Automatic Comparison

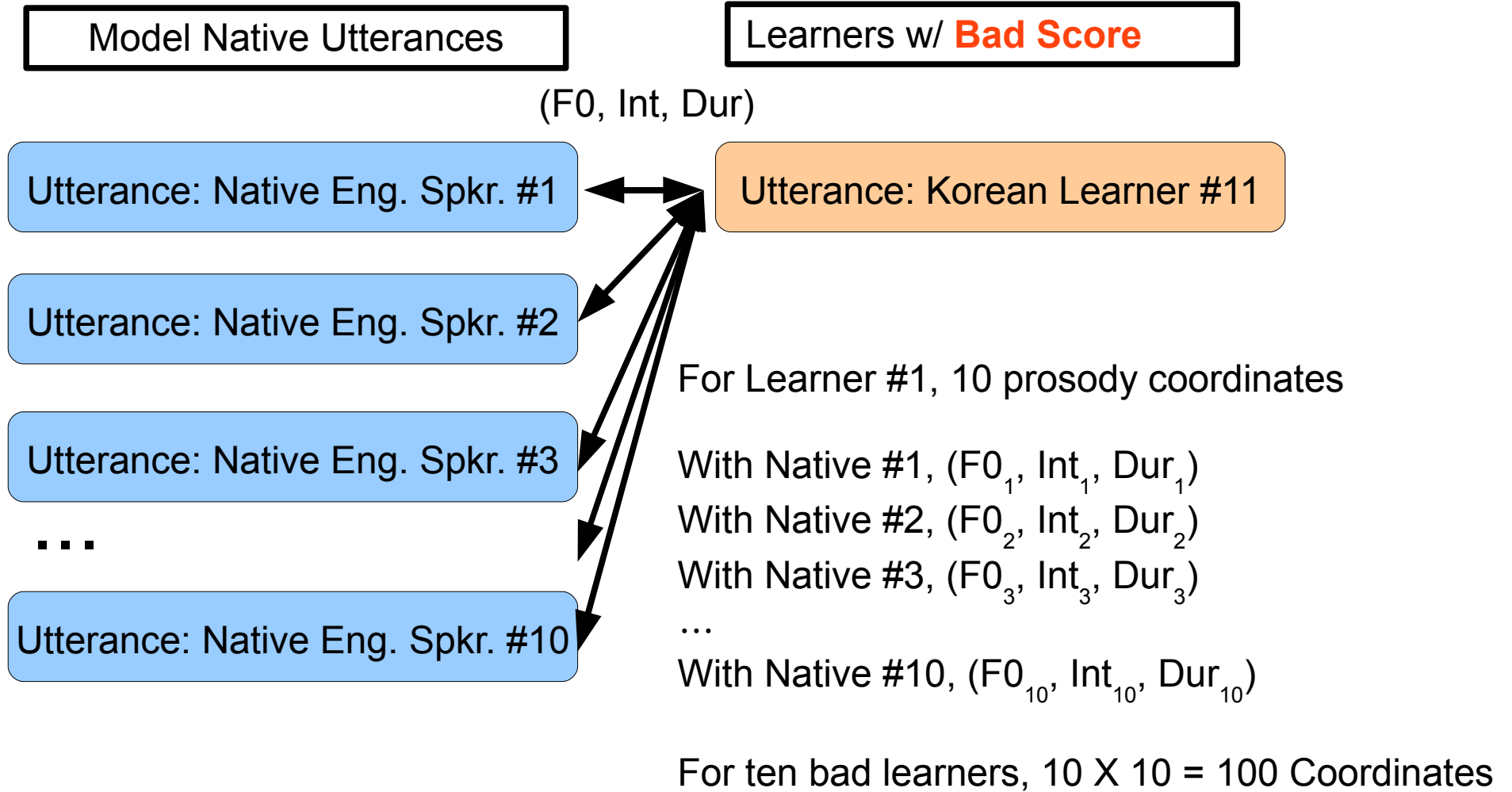


# 3D Space of Prosody Coordinates

Prosody coordinates of good learners

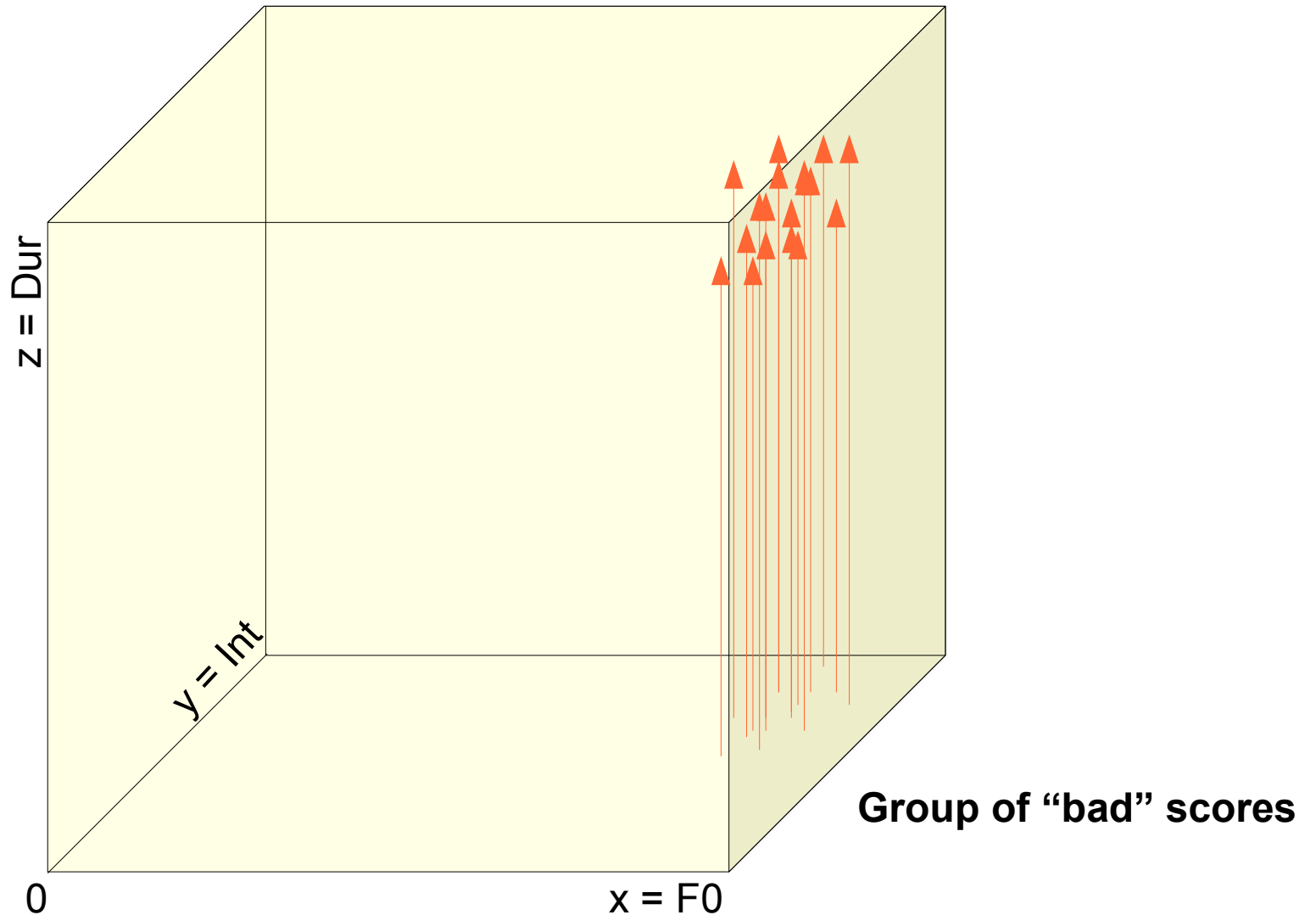


# Automatic Comparison

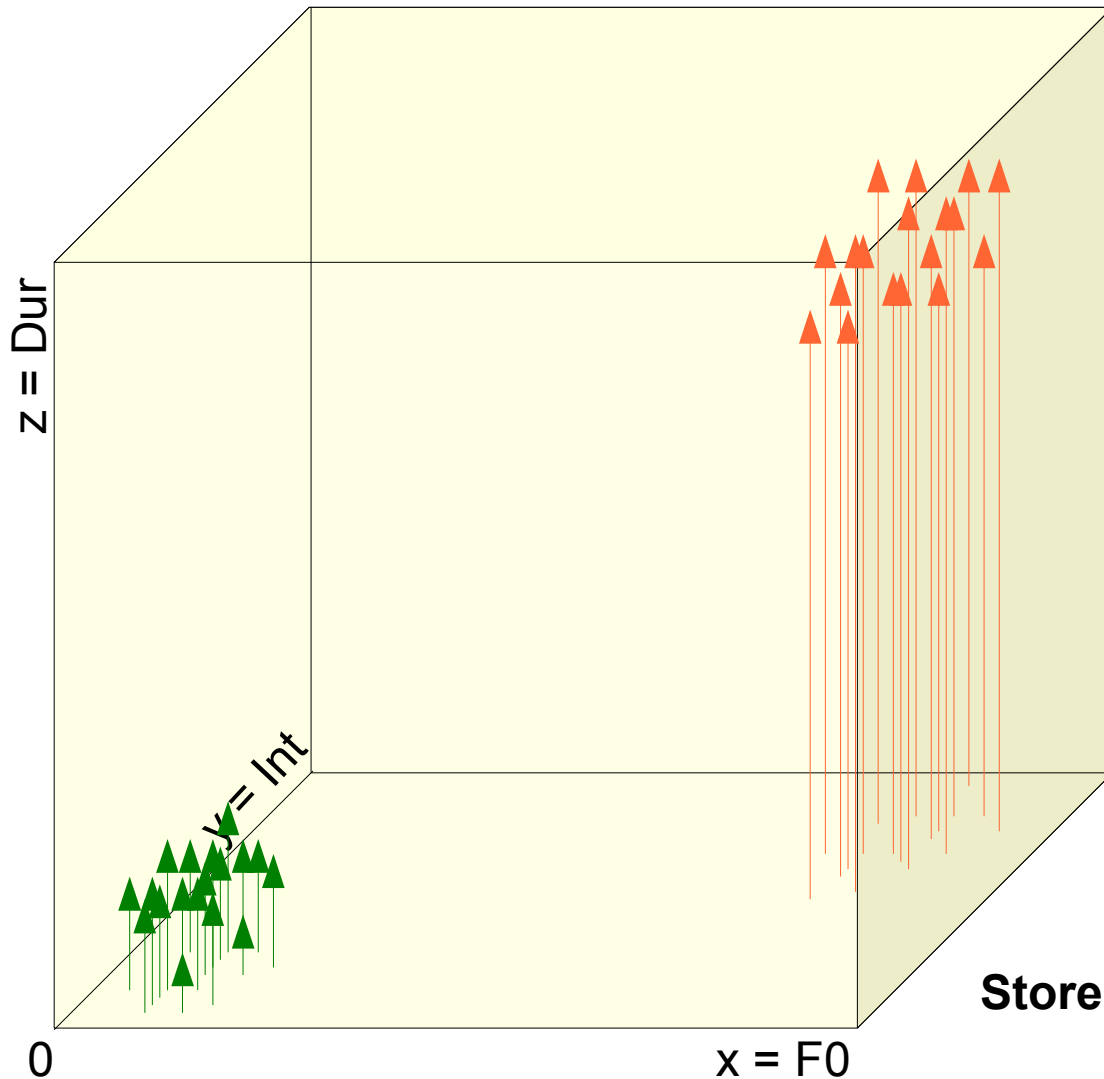


# 3D Space of Prosody Coordinates

Prosody coordinates of bad learners



# 3D Prosody Model – *an ideal case*



## Location of arrow heads

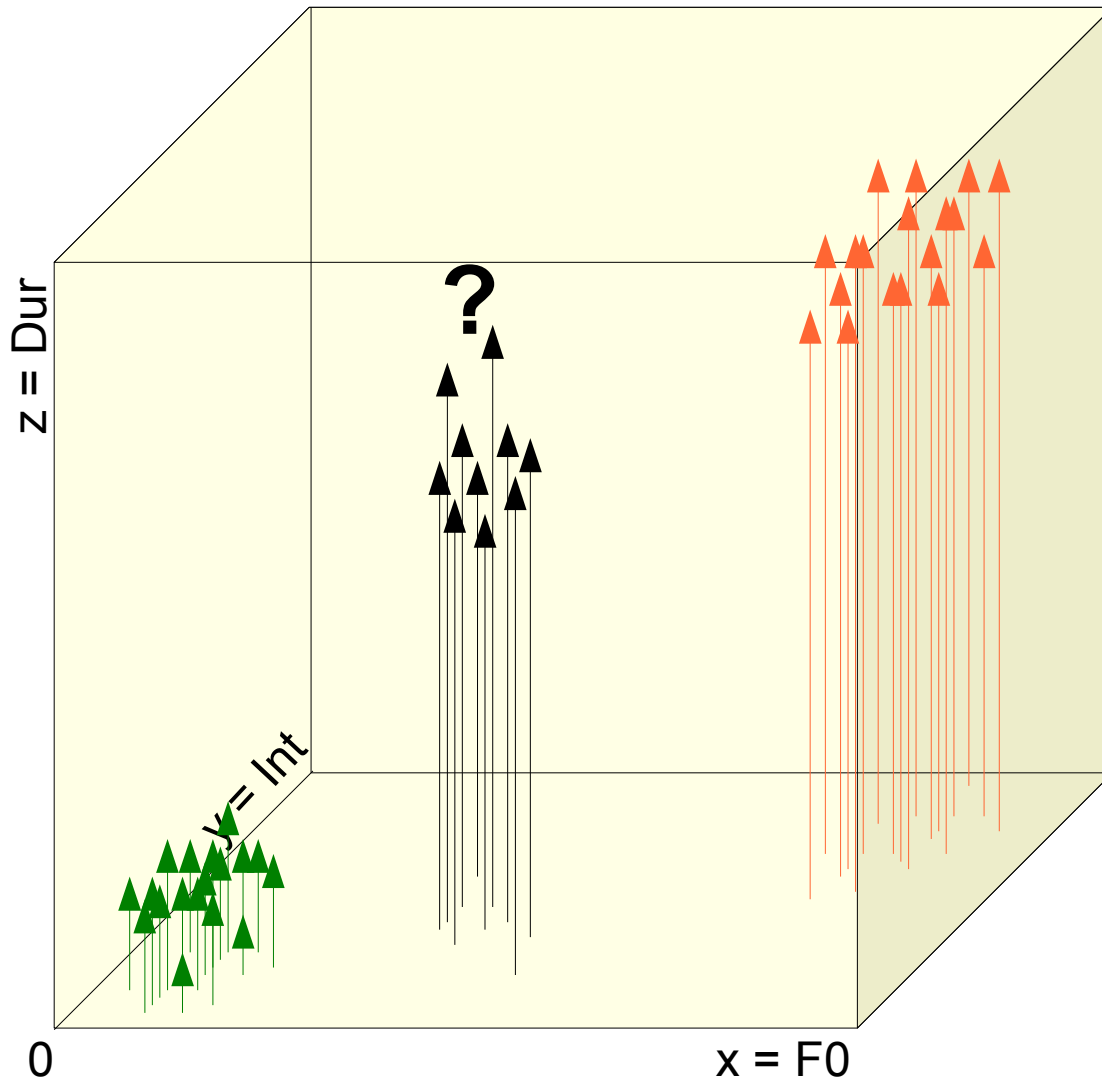
→  
Automatic comparison  
of learner utterances  
with native speakers'  
in the form of 3D coordinates,  
i.e.  $(x, y, z) = (F0, Int, Dur)$

## Color of arrow heads (Group membership)

→  
Human evaluation  
of learner utterances  
in terms of “good” and “bad”

**Store this model in a discriminant function**

# Discriminant Analysis



**A learner with no manual score,  
i.e. with unknown membership**

A set of automatically computed  
prosody comparison coordinates

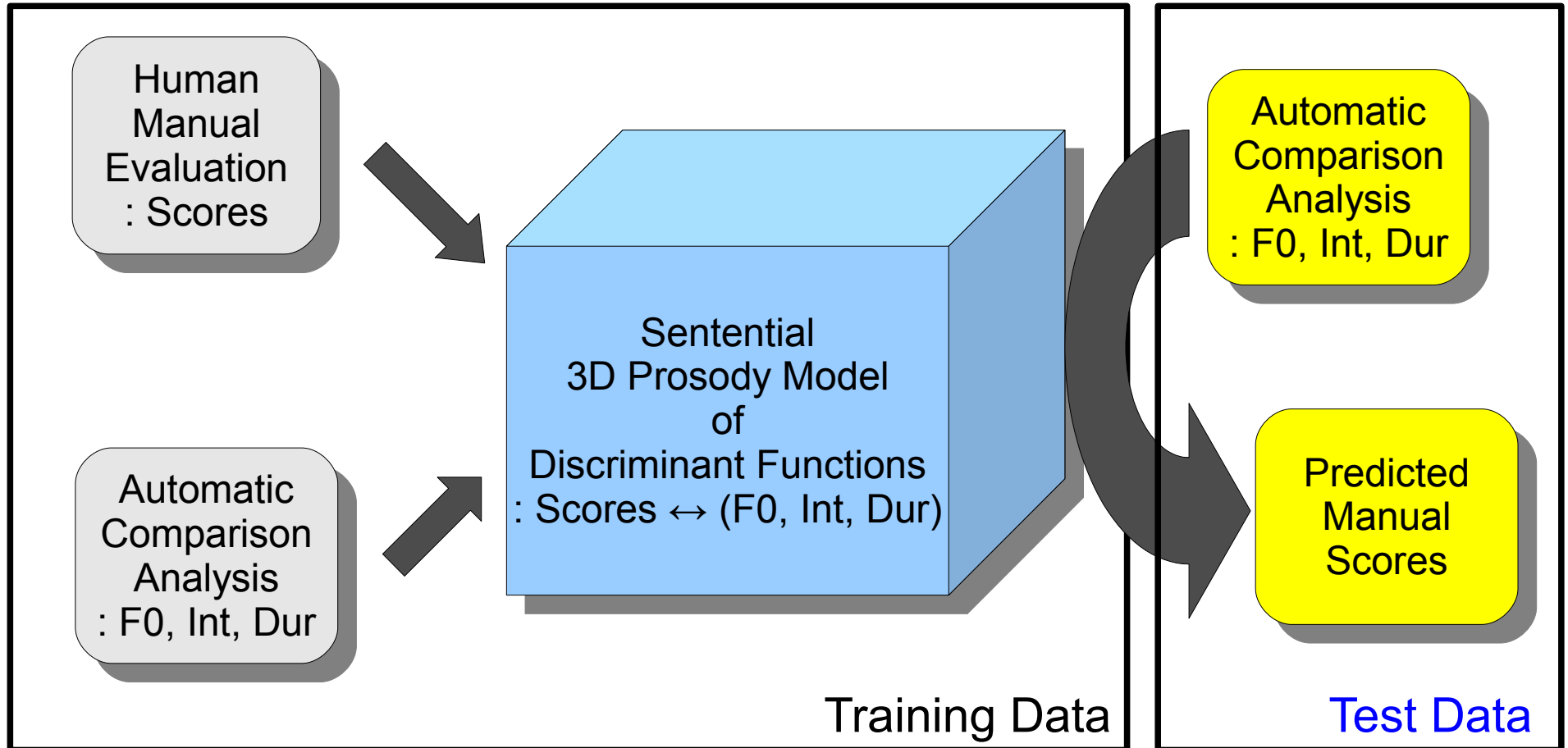


**Perform a discriminant analysis**



**Predict its group membership**

# Automatic Assessment of Prosody



# Human Evaluation

Manual prosody scores for  
 “The dancing queen likes only the apple pies”

Group	File	Segment	Overall	Group	File	Segment	Overall	Group	File	Segment	Overall
<b>N5</b>	File 1	5	5	<b>K5</b>	File 13	4	5	<b>K2</b>	File 21	4	3
	File 2	5	5		File 14	4	4		File 22	4	3
	File 3	5	5		File 15	4	4		File 23	4	3
	File 4	5	5		File 16	4	4		File 24	4	3
	File 5	5	5		File 17	4	4		File 25	4	3
	File 6	5	5		File 18	4	4		File 26	4	3
	File 7	5	5		File 19	4	4		File 27	4	3
	File 8	5	5		File 20	4	4		File 28	4	3
	File 9	5	5				File 29		4	2	
	File 10	5	5				File 30		4	2	
	File 11	5	5								
	File 12	5	5								

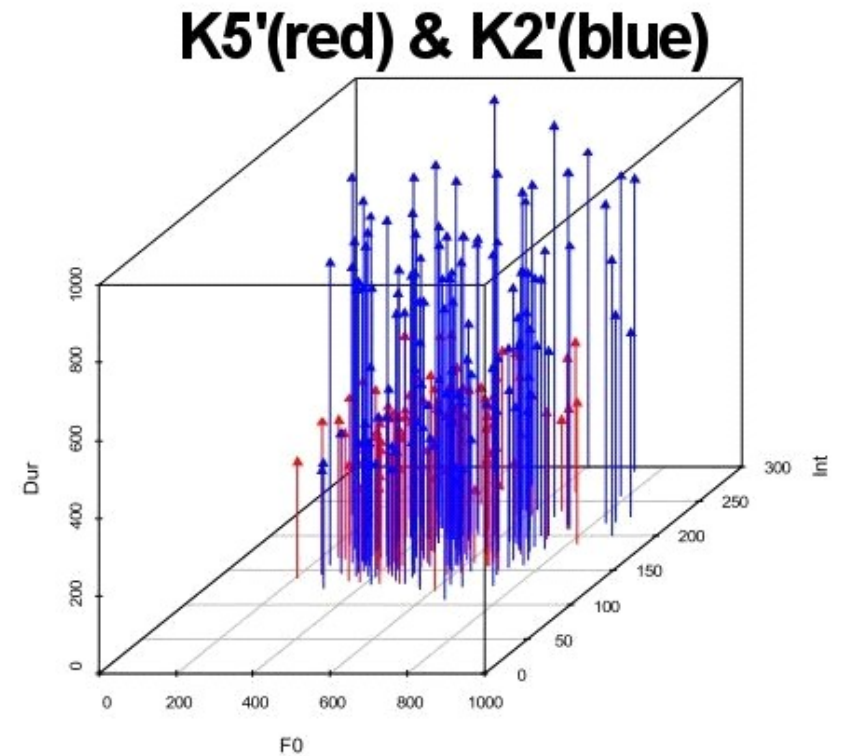
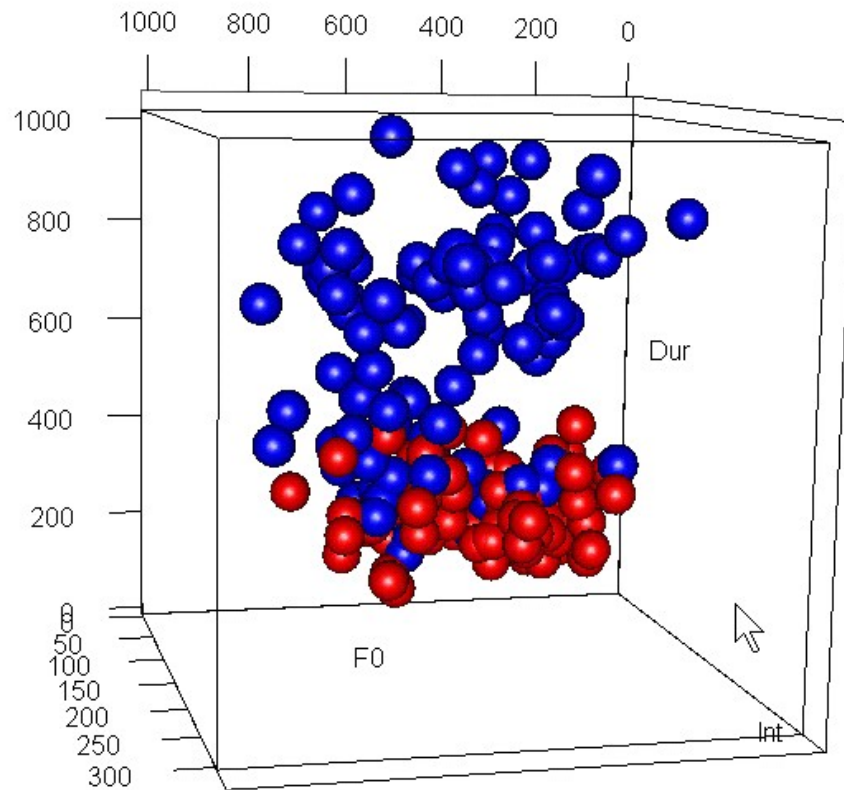
<Table 2> The segment and overall scores of the 30 utterances for Set B sentence “The dancing queen likes only the apple pies”. The evaluation was performed by a Korean phonetician. Note that the segment evaluation scores were controlled among the Korean learner groups K5' and K2'.

## Sample coordinates for one utterance from group K5

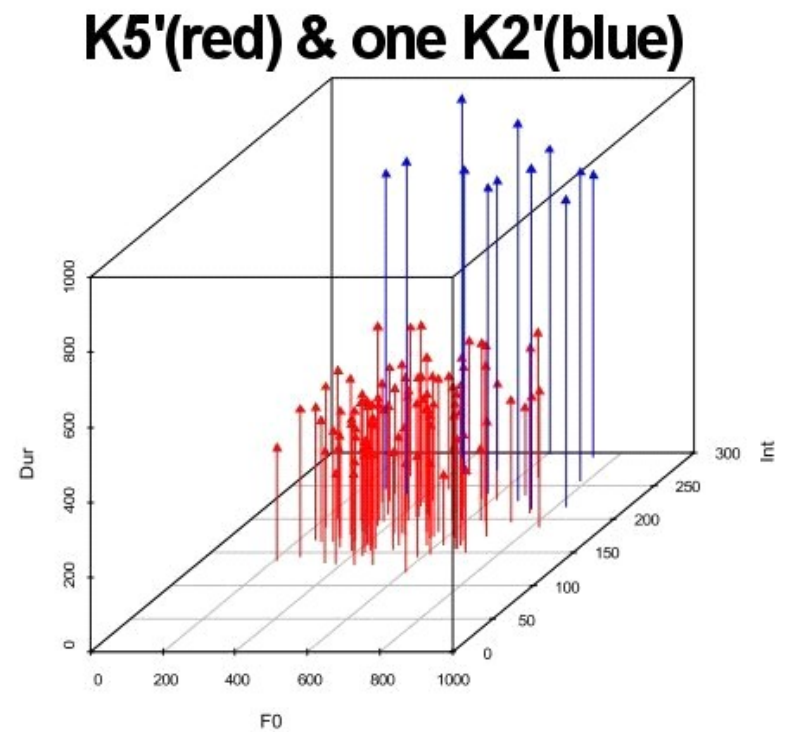
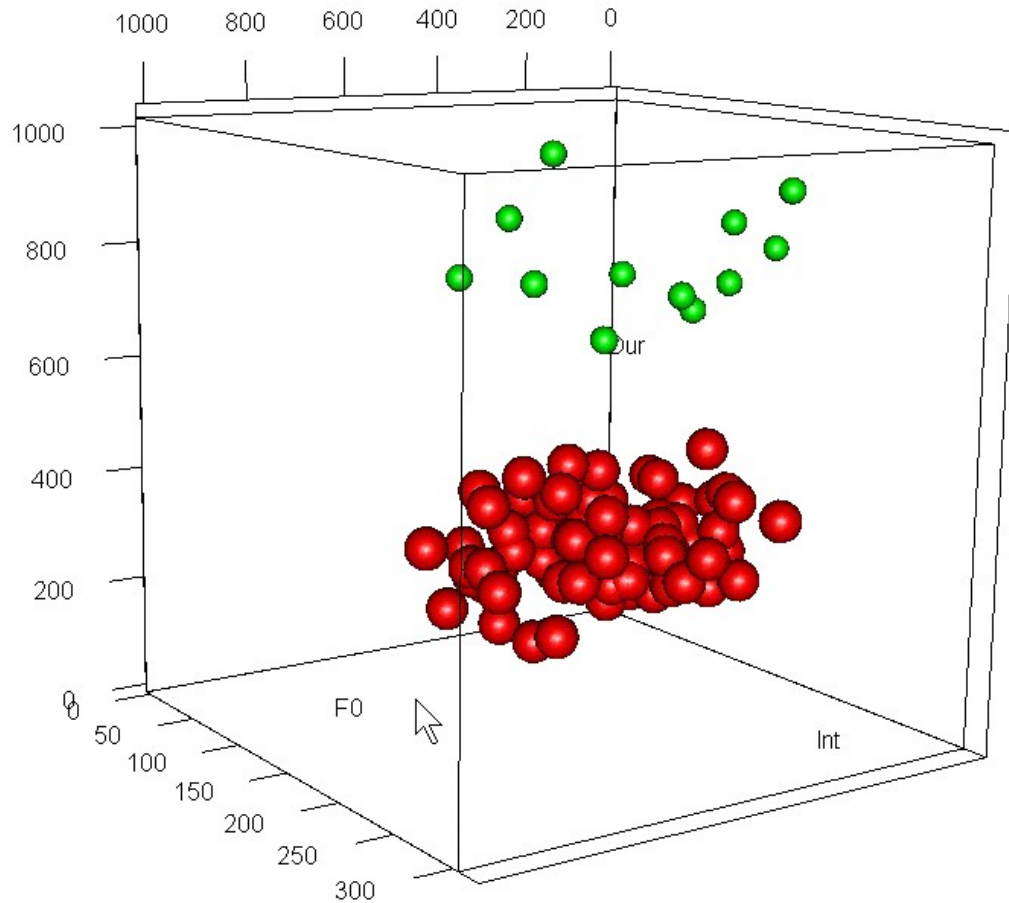
Comparison between	F0 contour difference	Intensity contour difference	Segmental durations difference	Coordinate (x, y, z)
N5.U1 and K5.U1	899	142	408	(899, 142, 408)
N5.U2 and K5.U1	360	92	190	(360, 92, 190)
N5.U3 and K5.U1	377	159	183	(377, 159, 183)
N5.U4 and K5.U1	206	81	151	(206, 81, 151)
N5.U5 and K5.U1	291	153	826	(291, 153, 826)
N5.U6 and K5.U1	251	113	563	(251, 113, 563)
N5.U7 and K5.U1	346	120	532	(346, 120, 532)
N5.U8 and K5.U1	299	114	343	(299, 114, 343)
N5.U9 and K5.U1	282	92	216	(282, 92, 216)
N5.U10 and K5.U1	716	178	183	(716, 178, 183)

<Table 3> A sample coordinates of the overall proficiency score points for K5U1 utterance. N5 and K5 represent the group names and U# represents the utterance number.

# 3D Model : K5 vs. K2



# 3D Model : K5 vs. one from K2



# Prediction with Discriminant Analysis

predicted observed	K2'	K5'	total probability	predicted observed	K2'	K5'	total probability
K5'	0.423926	0.576074	1.000000	K2'	0.973646	0.026354	1.000000
K5'	<u>0.586402</u>	0.413598	1.000000	K2'	0.997847	0.002153	1.000000
K5'	0.144321	0.855679	1.000000	K2'	0.995902	0.004098	1.000000
K5'	0.356046	0.643954	1.000000	K2'	0.994552	0.005448	1.000000
K5'	0.198097	0.801903	1.000000	K2'	0.977365	0.022635	1.000000
K5'	0.247591	0.752409	1.000000	K2'	0.985658	0.014342	1.000000
K5'	0.226376	0.773624	1.000000	K2'	0.986786	0.013214	1.000000
K5'	0.181814	0.818186	1.000000	K2'	0.980868	0.019132	1.000000
K5'	0.343508	0.656492	1.000000	K2'	0.986836	0.013164	1.000000
K5'	0.286556	0.713444	1.000000	K2'	0.988487	0.011513	1.000000
K5'	0.161533	0.838467	1.000000	K2'	0.985466	0.014534	1.000000
K5'	0.203144	0.796856	1.000000	K2'	0.981090	0.018910	1.000000

<Table 4> The classification table from the discriminant analysis. The number in each cell represents the probability of the automatic prosody score being classified into the predicted group. The left panel is for the test learner from K5' and the right panel is for the test learner from K2'.

predicted observed	K2'	K5'	total
K2'	12	0	12
K5'	<u>1</u>	11	12
total	13	11	24

<Table 5> The confusion matrix for the classification table.

# References

- [1] Boersma, Paul, “Praat, a system for doing phonetics by computer”, *Glott International* 5(9/10), pp.341-345, 2001.
- [2] Moulines, E. & F. Charpentier, “Pitch synchronous waveform processing techniques for text-to-speech synthesis using diphones”, *Speech Communication* 9, pp.453-467, 1990.
- [3] Rhee, S., S. Lee, Y. Lee & S. Kang, “Design and construction of Korean-Spoken English Corpus (K-SEC)”, *Malsori* 46, pp.159-174, 2003.
- [4] Yoon, K, “Imposing native speakers' prosody on non-native speakers' utterances: The technique of cloning prosody”, *Journal of the Modern British & American Language & Literature* 25(4), pp.197-215, 2007.
- [5] Yoon, K. “Synthesis and evaluation of prosodically exaggerated utterances.” *Phonetics and Speech Sciences* 1(3), pp.73-85, 2009.