A Cross-Language Comparison of Speaking Rate Effects on the Production and Perception of English Word-final Stops

Seok-Han Kang
Language Institute, University of Incheon
Email: kangseok@incheon.ac.kr

Abstract

The primary goal of this study is to find out how the effect of speaking rate has some influence on the production and perception across languages. Through both experiments of production and perception, an English native speaker changes both production and perception simultaneously. Especially the production of the temporal features changes relatively fast. On the contrary, Chinese and Korean speakers changes their production rather than perception by following the speaking rate.

I. Introduction

The purpose of this study is to understand how the effect of speaking rate affects the production and perception in English word-final stops across Chinese, Korean, and English native speakers. The effect of speaking rate has been well studied in English [1][2][3] as well as other languages [4]. Little research, however, has been documented on the English word-final stops cross -linguistically. The present study examined three questions: First, does speaking rate have comparable effects on the production and perception of word-final stops across various language users? Second, can we find the difference between production and perception universally? Finally, what is particular characteristics for each language group?

Of interest in this study was whether we could find the universal characteristics in both production and perception in spite of different speaking rates. That is, can we find the universal cues transcending languages, contexts, and speakers? In this study, 3 participants (each Chinese, Korean, and English native speaker) participated in both experiments. The criteria for producing separate rates derived Miller and Baer [1983], in which the criteria of “slow speaking rate”, “casual speaking rate”, and “fast speaking rate” were used.

II. Production experiment

In the production experiment, I measured the frequency features such as fundamental formant (f0), the first formant frequency, and the second formant frequency, as well as the temporal features such as the duration of preconsonantal vowel, closure, and release burst depending on three different rates. The minimal pairs and the frame sentence used in the present study were provided below:

(1) words: bap, bab, bat, bad, back, bag, pap, pab, pat, pad, pack, pag, dap, dab, dat, dad, dack, dag, tap, tab, tat, tad, tack, tag, gap, gab, gat, gad, gack, gag, cap, cab, cat, cad, cack, cag

frame sentence: ‘Say ____.’

All subjects were asked to read the words embedded in the frame sentence randomly ordered three times. Recording was done in a sound-proof broadcasting booth at the University of Incheon. Sounds were recorded on TASCAM Digital Recorder with a Shure unidirectional dynamic microphone.

Recorded sounds were digitalized at a sampling rate of 44,100 Hz and a quantization rate of 16 bits. The results were like follow tables.

Table 1. The mean of pre-consonantal vowel duration (ms), closure (ms), release burst (ms), fundamental frequency (Hz), the first formant (Hz), and the second formant (Hz) for an English native speaker in the word final English stops.

	pre-vowel		f0		F1		F2		closure		release	
	vd	vl	vd	vl	vd	vl	vd	vl	vd	vl	vd	vl
Slow	335	255	125	121	650	706	1549	1610	69	95	125	124
Casual	334	257	116	107	612	681	1660	1604	56	64	64	71
Fast	295	238	106	114	639	692	1665	1608	55	61	54	53

Note: Fundamental and formant frequencies are measured from the point of 100 ms where the glottal peak in the periodic portion of the voiced speech begins.

We found that the changing degree was clear for the temporal features rather than the frequency features in the case of an English native speaker. Along with the decrease of the speaking rate, the duration of the pre-consonantal vowels decreased as 11% in voiced sounds and 6 % in voiceless sounds, the closure duration as 20% in voiced and 35% in voiceless, release as 57% in both sounds. On the contrary, the changing degree of the frequency features were small relatively. Fundamental frequency decreased as 15% in voiced sounds and 5% in voiceless sounds, and the first formant as 2% in both sounds. We couldn't analyze the second formant because of the minimal difference.

Table 2. The mean of pre-consonantal vowel duration (ms), closure (ms), release burst (ms), fundamental frequency (Hz), the first formant (Hz), and the second formant (Hz) for a Chinese native speaker in the word final English stops.

	pre-vowel		f0		F1		F2		closure		release	
	vd	vl	vd	vl	vd	vl	vd	vl	vd	vl	vd	vl
Slow	221	172	113	113	585	585	1820	1887	42	99	77	126
Casual	212	151	113	111	587	571	1748	1785	49	82	79	120
Fast	182	149	105	103	570	548	1805	1812	46	81	82	118

In the case of a Chinese native speaker, the changing degree was clear only for the pre-consonantal duration. Along with the decrease of the speaking rate, the duration of the pre-consonantal vowels decreased as 18% in voiced sounds and 13 % in voiceless sounds. On the contrary, the changing degree of the closure duration and the release burst was small relatively. Fundamental frequency decreased as 7% in voiced sounds and 8% in voiceless sounds, and the first formant as 9% in voiced sounds and 6% in voiceless sounds.

Table 3. The mean of pre-consonantal vowel duration (ms), closure (ms), release burst (ms), fundamental frequency (Hz), the first formant (Hz), and the second formant (Hz) for a Korean native speaker in the word final English stops.

	pre-vowel		f0		F1		F2		closure		release	
	vd	vl	vd	vl	vd	vl	vd	vl	vd	vl	vd	vl
Slow	217	186	119	122	615	630	1812	1765	80	102	28	81
Casual	199	169	104	101	613	626	1821	1769	78	75	36	65
Fast	198	174	106	100	618	623	1829	1789	75	84	33	70

In the case of a Korean native speaker, the changing degree was clear relatively for the pre-consonantal duration, and fundamental frequency, and the closure duration. Along with the decrease of the speaking rate, the duration of the pre-consonantal vowels decreased as 9% in voiced sounds and 6 % in voiceless sound, fundamental frequency as 11% in voiced sounds and 18% in voiceless sounds, and the closure duration as 6% in voiced sounds and 18% in voiceless sounds. On the contrary, the changing degree of the first and the second formant and the release burst duration was small relatively.

III. Perception experiment

A highly educated native speaker of American English was recruited to make stimuli. He was a 24-year-old male who was born and raised in Michigan. He had no history of speech disorders and no problems in pronunciation. Minimal pairs of English stops were selected across places of articulation (bilabial, alveolar, and velar) in the final position.

Each unit segment was encoded as voiced or voiceless. Then I produced a sequence of 2 digit signals (the preceding vowel and the closure duration in non-release condition) or 3 digit signals (the preceding vowel, the closure duration and release duration in release condition) by repeated permutation, where each interval functions as a binary digit of 0 and 1.

The ExperimentMFC program on Praat was used for the perception tests. Each stimulus was randomly ordered and repeated three times. The perception tests were conducted in a quiet office. Subjects who already took part in the production experiment took perception test after a week. They were asked to click the mouse on one of the two words presented on the screen wearing a headset. It took around 30 minutes for each subject to complete the series of tasks. The subjects' responses were statistically analyzed with SPSS 10.0.

Table 4. Result of Cue Identification Test (%)

	word final release			word final non-release		
	slow	casual	fast	slow	casual	fast
English	100	100	95	95	100	100
Chinese	100	95	73	89	89	62
Korean	89	95	100	78	78	95

The perceptual experiment was to examine the subjects' accuracy for consonant identification in word final position depending on the presence or absence of the release-burst. English native speakers identified the pure stop voicing signals more clearly than Chinese and Korean subject in all three rates. As shown in <Table 4>, Chinese made more errors perceptually in non-released stops, especially in the fast speaking rate.

IV. Result

On the whole, an English native speaker changes both production and perception along with the speaking rate. Especially the production of the temporal features changes relatively fast. On the contrary, Chinese and Korean speakers changes their production, not the perception by following the speaking rate. A Chinese speaker's production on the duration of the pre-consonantal duration changes the most along with the speaking rate, while perceptually the cue of the release burst is meaningful in any speaking rates. It means that they get the low score without the release burst. A Korean speaker shows the relative change in the pre-consonantal vowel and the fundamental frequency, while the cue of the release burst is meaningful only in the slow speaking rate.

References

[1] Miller, J. L., & Baer, T. Some effects of speaking rate on the production of /b/ and /w/. Journal of the Acoustical Society, 73, 1751-1755. 1983.

[2] Miller, J. D. Auditory-perceptual interpretation of the vowel. Journal of the Acoustical Society of America, 85, 2114-2134. 1989.

[3] Jae-Jin Sa, Effects of Speaking Rate on Segmental Duration in English. Ph.D. Dissertation. Hankook University of Foreign Studies. 2005.

[4] Schmidt M. & Flege, E. Speaking rate effects on stops produced by Spanish and English monolinguals and Spanish/English bilinguals. Phonetica. 53. 162-179. 1996.

[5] Miller, J. L., & Baer, T. Some effects of speaking rate on the production of /b/ and /w/. Journal of the Acoustical Society, 73, 1751-1755. 1983.