A study of asymmetry relations between perception and production in the word-final stop voicing*

Seokhan Kang
(Konkuk University)

Kang, Seokhan. 2014. A study of asymmetry relations between perception and production in the word-final stop voicing. *Studies in Phonetics, Phonology and Morphology*. 20.1. 3-22. This study tries to investigate whether the main acoustic properties in the production differ from those in the perception. The experimental design is intended to support the hypothesis that phonetic cues with equal amounts of acoustic information by production should be reinterpreted in the perception level. In the experiments, how the real prominent cues influence perceiving English stop voicing is examined. The results from both experiments clearly prove the asymmetry between production and perception. Even though phonetic cues measured in the production exert similar statistical significance, they could be re-ordered or re-arranged in the level of perception. Through the experiments, this study supports the weak or indirect theory between the two. **(Konkuk University)**

Keywords: Production, perception, voicing, phonology, phonetics, asymmetry

1. Introduction

The identification of the acoustic cues to the voicing contrast between minimal pairs of stops has been an issue for a long time in both production and perception. These acoustic cues which influence voicing contrast include VOT, F0 perturbation, the presence or absence of phonation during closure, vowel duration, closure duration, consonant/vowel duration ratios, the presence of the release burst, the duration of the release burst, F1 transition, and F2 transition, etc.

Some research points out that these phonetic properties play different roles in realizing stop phonemes (Plauché 2001), in which VOT is suggested to be the most distinctive cue in classifying laryngeal features (Iverson and Salmons 1995) as well as in voicing (Lieberman et al. 1958, Lisker and Abramson 1964). However, other researchers raise questions on whether the cue of VOT is represented as the most crucial cue even in perceiving the stop voicing (e.g., Williams 1977, Serniclaes and Bejster 1979, Nearey 1997). Doubts arise from the results of the perceptual experiments which use manipulation of natural phonemes (Williams 1977) or artificial sounds (Serniclaes and Bejster 1979). They report that other properties may replace VOT under some conditions. For example, Kang (2005) suggests that the vowel duration replaces the role of VOT in perceiving the English stop voicing in the trochaic environment. It means that the perceptual strength of VOT could be reinterpreted differently if prosody adds to the segmental

* This work was supported by Konkuk University.

perception.

The recent studies reach the conclusion that cue-trading effect could take place when other factors are involved in the decision process. Interestingly it is widely agreed that this effect has serious influence on the relationship between production and perception. Borden and colleagues (1994) argue how the listeners process the phonetic signals and elicit the acoustic information in perception is not still revealed because of opaque relations between the two. Plauché (2001) reports that how the phonetic cues process the auditory system has not been investigated yet. These problems result from physical variations inherently resided in phonetics; the cues in the acoustic signals could realize differently in each context and have uneven influence on phoneme contrast. These variations cause the asymmetry relations between phonetic cues (from variable production) and the phoneme features (from invariable perception).

Because of comparatively rare investigation on the relationship between auditory system and phonetic cues, some research on phonetic-based phonology tends to rely more on amounts of phonetics cues which presuppose the cue-weight. For instance, in the theory of 'licensing by cue' (Steriade 1995, Flemming 1995), the amounts of acoustic information are measured, centering on the numbers of phonetic cues. Thus, their cue weighting is supposed to be decided by the acoustic features in the production level. In reality, it is problematic that the dependency on the production may distort the real perceptual weight.

This study tries to investigate whether the main acoustic properties in the production differ from those in the perception. That is, phonetic cues with the equal amounts of acoustic information by production may be reordered in the perception level. Thus, the study would try to look for the real prominent cues when the native English speakers perceive the voicing of English plosives.

If this study proves that the perception is more appropriate in reporting the cue weighting than the production, this study would supplement the weak point of phonetic-based phonology theories (e.g., theory of Licensing by Cue) which propose the phoneme contrast/neutralization mainly by the amounts of phonetic information. Thus, this study would try to support the hypothesis that the cue weight could be decided mainly by perceptual quality, not by the amounts of acoustic information. More specifically, it is suggested that the asymmetry between production and perception should be considered in the phonetic-based phonological theories.

2. Phonetic voicing cues in word-final position

The essential articulating feature of a stop consonant is a momentary blockade of the vocal tract and a successive release of the oppressed air. Naturally, the stops require a total blockade between active and passive articulators. The stops found in the universal languages generally go through

process of approach, closure, and release. The voicing contrast for the English stops could be verified by realizing the cues in each environment (Kingston and Diehl 1994).

Table 1. The voicing contrast in English plosives

	[+voiced]	[-voiced]
VC	long pre-consonantal vowel duration	short pre-consonantal vowel duration
	voicing during closure	silence during closure
	short stop closure	long stop closure
	low F1	high F1

In English, voiced stops often have no vocal fold activity in syllable final position. This means that the listener must rely on other cues to perceive stop voicing. We can find various voicing cues in word-final position: pre-consonantal vowel duration, F1, F2, fundamental frequency (F0), closure duration, release duration, or amplitude.

3. Theoretical background

Previous research creates various theories on how phonetic cues affect the stop voicing from the articulation, perception, and auditory (e.g., Lieberman and Mattingly 1985, Kingston and Diehl 1994, Stevens 1999, Hayward 2000). The theories have been set up focusing on whether its relationship is unidirectional or bidirectional, and strong or weak.

Strong relation theory refers to a simple, robust, and transparent relation between articulation and auditory, or between production and perception. Its types include the double-strong theory (Stevens 1999, Hayward 2000), strong gestural theory (Lieberman and Mattingly 1985), and strong auditory theory (Kingston and Diehl 1994). Double-strong theory implies the strong and clear relations between articulation and auditory, or production and perception. Strong gestural hypothesis postulates a strong connection between production and perception, but there is a more complex and unidirectional path between articulation and auditory properties which reflect the nonlinear relations. Strong auditory theory assumes that there is a strong relation between phonology and auditory properties but only weak and indirect between phonology and gestures. Regardless of their different perspectives, strong theory is thought of a strong/direct relation between perception and production.

The criticisms on the strong relation theories postulate that there is only weak and opaque relation because perception may dominate over other areas. Nearey (1997) argues that speakers optimize their phonetic realization by

minimizing articulated effort and maximizing the perceptual distinctiveness. Flemming (1995) supports the theoretical arguments that perception shapes production, not in reverse order. Kang (2005) suggests that perceptual constraints arrange both sound inventories and phonological process in languages. These studies strongly imply the weak or indirect relations.

The weak theory suggests an indirect mapping between production and perception. Unlike the strong theory, the relationships are not necessarily straightforward. They are assumed only to be tractably systematic, to posit within the feasible range of both systems. The suggestion is that the primary source of phonological patterns results from the perceptual properties. Such properties which are plausibly involved in the case include the ratio of vowel and consonant in a syllable and the low-frequency property which powerfully influences human auditory system. In this respect, a weak approach assumes temporally distributed relational features which can be incorporated into localized phonological distinctions.

Again, a weak hypothesis may admit indirect and more complex relations between phonological elements and physical properties than other approaches. It allows considerable flexibility in the values of phonetic cue information among segmental/suprasegmental features. From this perspective, both experiments of production and perception should be examined simultaneously to prove the hypothesis.

4. Production

4.1 Experiment

The data were collected from 10 adult native English speakers. At the time of recording, they were visiting students who learned Korean at a Korean Language Institute attached to the university. Most of the recordings had been carried out in a university in Seoul. None reported being diagnosed with a language or speech disorder. They had arrived in Korea at 23.0 years old (Standard Deviation = 3.2) and had stayed in Korea for 4.93 years (SD = 2.2).

Minimal pairs of cognate English plosives were selected from different places of articulation (bilabial, alveolar, and velar) in word-final position. The minimal pairs were embedded in frame sentences to minimize the effects of the other factors. The minimal pairs and frame sentences used in the present study are provided below:

(1) Frame sentence

Say _____. **:** ca**p** ca**b** ca**t** ca**d** ca**ck** ca**g**

The total numbers of analyzed words were 288 (8 speakers * 2 voicing * 3

articulate places * 6 words)[†]. The selected words were under the same condition except for the voicing contrast of cognate stops. Each participant was asked to read each English sentence three times. Before they produced the sentences, it was confirmed that they knew what the sentences meant, and that they knew how to pronounce them. The sounds were recorded with a Marantz PMD 650 using a Shure SM 10A microphone, digitalized at 44.05 kHz and 16 bit resolution.

The results of the acoustic measurements for the stop voicing obtained were compared. Several acoustic measurements in fundamental frequency (in Hertz) and duration (in milliseconds) in consonantal and feature alternation as well as in the first and second formants (in Hertz) were made. These acoustic cues were measured using a waveform display with a time-locked wideband spectrogram with the software Praat (5.1.25). All acoustic cues were measured from the initial acoustic signal in both the waveform and the spectrogram to the final acoustic cues of the boundary such as burst or spectral cues (Kent and Read 2002). These measures were analyzed with independent two way ANOVAS which was conducted for statistical evaluation with the following parameters: dependent variables of the duration of the pre-consonantal vowels, closure, and the release burst, and independent variables of environment and voicing.

4.2 Result

The statistical results for all cues measured are as follows;

Table 2. The statistical results of the two-way ANOVA

	variables	main effects				cohort effects	
		Voicing		articulatory place			
		F(1,287)	P	F(2,287)	P	F(2,287)	P
word-final released	pre-consonantal vowel du.	135.922	.000	4.877	.008	1.673	.190
	closure	3.701	.056	7.313	.001	4.584	.011
	release	.076	.784	19.749	.000	.667	.514
unreleased	post-consonantal vowel du.	9.331	.008	1.936	.177	.065	.937

The statistical measurement reports that release is not the meaningful cue in defining the voicing ($p > .05$). In this study, 168 out of 288 (59%) produce the release. Among the released stops, the weak released words with the

[†] 2 speakers were excluded because of recording problems.

duration less than 10 msec occupy 44 % (73 words), along with the medium released words between 11ms to 30ms as 23% (38 words) and the strong-released words over 31 msec as 34% (57 words).

In case of the released words, there is no difference in the voicing: 50.4% for the voiceless stops and 49.6% for the voiced stops. For the articulatory places, the frequency of the release can be found in the velar (66%), alveolar (55%), and bilabial (48%). Byrd (1993), analyzing 54,000 words in TIMIT data, reports that release words are more frequent than the non-released words as 59.7% vs. 40.3%. She concludes that there is difference not in the voicing, but in the places ($p < .0001$): the bilabial stops (49.5%), the alveolar stops (57%), and the velar stops (83.1%). The results imply that the frequency of the release could be fewer in the non-laboratory conversation which TIMIT database collected and also the criteria (on what is the release) might influence the results. It could be possible that more strict criteria may cause the fewer frequency of the release.

In spite of the variation, the research points out that there is little difference in the voicing, while the difference can be found in the articulatory places; the velars are more frequent than the bilabial and alveolar stops. It means that the air pressure in the back of the oral cavity may produce stronger release. Table 3 presents the mean values and standard deviation of pre-consonantal vowel, closure duration, and release in word-final released environment; and pre-consonantal vowels in word-final unreleased environment.

Table 3. The mean and standard deviation of voiceless stop cues for word-final environment

		Word-final released		Word-final unreleased	
		Mean	s.d.	Mean	s.d.
Closure duration	voiceless	84	39		
	voiced	74	47		
Release	voiceless	30	28		
	voiced	30	30		
Pre-consonantal vowel	voiceless	171	33	181	38
	voiced	224	40	234	28

*unit: msec

As a whole, the duration of the word-final syllable is presented as 307 to 308 msec. In details, the voiced stops are longer than the voiceless stops as 329 msec vs. 287 msec, in which the difference mainly emerges from the duration of the pre-consonantal vowels. However, there is no difference in the places: 308 msec for the bilabials, 307 msec for the alveolar stops, and

306 msec for the velar stops.

The duration of the release is the longest for the velar stops. But, the duration of the preceding vowels is the longest for the alveolar stops, along with the closure duration for the bilabials. For instance, the duration of the release for the velar stops is two times as long as that of other stops. Also the duration of the preceding vowels for the voiceless alveolar stops (184 msec) is longer than that of other stops (around 15 to 20 msec). For the preceding vowels, the voiced stops show longer duration than the voiceless stops as 171 msec vs. 224 msec, while the alveolar stops are the longest as 207 msec. The voiceless stops are longer than the voiced one as 10 msec in the closure duration, showing 84ms vs. 74ms. In the articulatory places, it was measured as the bilabials (92 msec), alveolar stops (77 msec), and velars (69 msec).

The duration of the release is not related with the voicing statistically ($p > .05$). However, it is closely tied with the articulating place ($p < .001$); the velars (as 44ms) are the longest of other stops (e.g., 23 msec of the bilabials and 22 msec of the alveolar stops). In case of the velar stops, there is a difference in the voicing: 42 ms of the voiceless stops and 48 ms of the voiced stops. Post-hoc Scheffe test for the voicing conducts as follows;

Table 4. Post-hoc test for VC environment

	Voicing
Pre-consonantal vowel duration	voiced > voiceless
Closure duration	voiced < voiceless
Release	voiced = voiceless

In short summary, for the pre-consonantal vowels in the word-final condition, the voiced stops show longer duration than the voiceless counterpart: 224 msec vs. 171 msec. Also the difference occurs in the closure duration as 74 msec (voiced) and 84 msec (voiceless). However, there is no difference in release burst as 30 msec for both voiced and voiceless stops. It means that the release is not a significant cue to classify the voicing from the production experiment.

5. Perception

The perceptual identification of the acoustic cues to the voicing contrast has been carried out. The study intends to support the hypothesis that the cues leading to the voicing distinction for English plosives are flexible and realizes differently depending on the context and environment. In word-final positions, the perceptual salience for the duration of the word-final stop closure depends on the presence or absence of the final release, in case that the release burst is available perceptually when the final release is produced.

Perception experiment is designed to investigate how the subjects respond

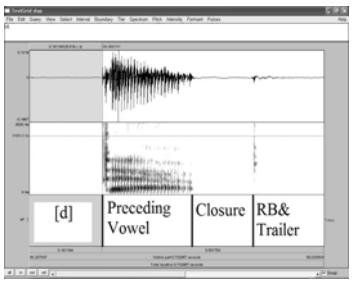
on the manipulated cues. The main purpose of the perception experiment is to explain the statistical meanings of the cues which measured in the production experiment. Specifically the goal encompasses two primary objectives. First of all, it checks how the native English subjects respond to the synthesized cues in perceiving the stop voicing. If they respond differently on the voicing depending on the cue sequences, it means that it proves the difference of the cue quality. Next, the experiments for both the response-match test and the cue-robustness test inform us what manipulated signals have more sensitivity. The response-match test tells us how the targeted cues influence the voicing, checking the match degree between synthesized cues and the original cues of stop voicing.

Another objective for the experiments is to test how cue quality forms, conducting the cue-robustness test. The test is designed to investigate how the targeted cue leads to the specific response, not masked by the neighbor cues, and how the targeted cues keep the perceptual robustness in perceiving the voicing contrast. Finally, the ultimate goal of the experiment proves the hypothesis of asymmetry relations between production and perception.

5.1 Methodology

The sound files used in the perception test emerge from two American English speakers who recruited to obtain minimal pairs of cognate stops to make stimuli for the present study. Both speakers were 30s' males who were born and raised in Illinois. None of them had any history of speech disorders.

Ten native speakers of English native speakers participated in the perception tests. The perception tests were conducted in the sound-proof laboratory immediately after the production test. One subject out of ten subjects was removed because of the outlying results. Most of the participants were exchanging students of a university in Seoul who had 0.7 years of residence in Korean and with mean ages of 22.1. Small compensation money was applied. Speech samples for the stimuli were extracted from the recorded sentences that showed segmental cues and phonological processes of the native speaker's recording.



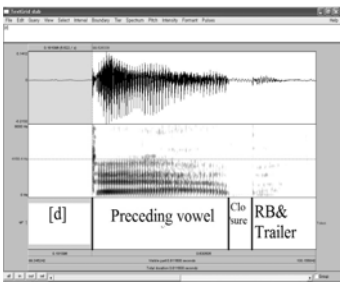


Figure 1. Segmentation of 'dap' **Segmentation of 'dab'**

The extracted speech samples were segmented by drawing a boundary between outstanding acoustic landmarks at a nearest zero-crossing point on Praat (Boersma and Weenink 2010). The durations of the segments are provided in Table 5.

Table 5. The composition of manipulating cues

	Files	Cue manipulation	Numbers
Word-final released	(1)111 (2)101 (3)110 (4)100 (5)011 (6)010 (7)001 (8)000	(1) voice V1 + voice closure + voice release (2) voice V1 + voiceless closure + voice release (3) voice V1 + voice closure + voiceless release (4) voice V1 + voiceless closure + voiceless release (5) voiceless V1 + voice closure + voice release (6) voiceless V1 + voice closure + voiceless release (7) voiceless V1 + voiceless closure + voice release (8) voiceless V1 + voiceless closure + voiceless release	8*3(places) = 24
Word-final unreleased	(1)11 (2)10 (3)01 (4)00	(1) voice V1 + voice closure (2) voice V1 + voiceless closure (3) voiceless V1 + voice closure (4) voiceless V1 + voiceless closure	4*3(places) = 12

Total stimuli: 36 samples (1 test) * 2 rounds (repeat) * 9 (participants) = 648

Examples of the segmentation of the unit intervals in word-final positions are illustrated in Figure 1 and Table 5. The label of RB and Trailer represents the release burst and its trailing signal. Minimal pairs of cognate stops are divided into a set of 3 unit intervals: the pre-consonantal vowel, the stop closure, and the release burst. Stimuli are made by combining the unit intervals. The order of the unit intervals is fixed, so that the release burst never precedes the stop closure. Each unit interval is labeled as 0 or 1 according to its voicing identity, such that 0 stands for 'an interval which is extracted from the word of a voiceless plosive' and 1 for 'an interval which is extracted from the word of a voiced plosive' Combination of the unit intervals labeled 0 and 1 provides bit strings of a length of four digits each of which allows the occurrence of 0 or 1. Thus, 000 stands for the original word including a voiceless plosive in the word-initial position, and 111 for the original word including a voiced plosive in the word-final position. 010 is a stimulus made of a sequence of the stop closure (0: voiceless), the release

burst and aspiration (1: voiced), and the following vowel (0: voiceless) in the word-initial position. As a result, 8 stimuli for word-final positions (000, 001, ..., 110, 111) are composed. Yet another set of stimuli (00, 01, 10, 11) is additionally made for the word-final unreleased condition by simply removing the final release to see how subjects respond to the stimuli lacking the final release. Therefore, the total number of stimuli amounts to 36. The manipulated sounds are presented by using Alvin program (Hillenbrand and Gayvert 2004) for the perception experiment.

5.2 The procedure of the experiment.

The cue editing is conducted in each environment using Praat and carried out by using Alvin continuously. 3 answering sheets are divided into each environment and total items are 108, repeating 2 times. The sound loudness is set at 60 dB, and designs to select answers forcefully. The mean times of test take around 20 minutes.

This perception experiment chooses the identification test. Generally the perception experiment can be divided into two: discrimination test which discriminate sounds for the given pairs of words and identification test which identify the sounds after hearing a series of words. In this identification test, the subjects are forced to identify the voicing, permitting the prediction. Thus, the study is classified into forced choice test because limited answers should be selected. Alvin program designed for the psychological purpose has various functions. For the experiment, the manipulated sound files edited by Praat load into Alvin by using play directory. The experiment is conducted by using laptop computer in the sound-proof lab. The distance between the participants and the computer is around 50 cm with 60 dB of the loudness.

5.3 Results

Following is the results of the experiment.

Table 6. The result of the perception test

	Files	voiceless response		voiced response	
		numbers	percentage	numbers	percentage
Word-final released	000	52	**96**	2	4
	001	34	**63**	20	37
	010	42	**78**	12	22
	011	49	**91**	5	9
	100	19	35	35	**65**
	101	15	28	39	**72**
	110	32	**59**	22	41
	111	1	2	53	**98**
Word-final	00	45	**83**	9	17

unreleased	01	44	**81**	10	19
	10	19	35	35	**65**
	11	8	15	46	**85**

*unit: percentage

In the file codes, 0 refers to the voiceless cue, while 1 refers to the voiced cue. For instance, the file of '110' in the word-final released environment consists of 'voiced vowel + voiced closure + voiceless release' cues. Also the file of '011' composes a series of 'voiceless vowel + voiced closure + voiced release'. The result is computed as the percentage and numbers of hit stimuli for statistical meanings. The results report that 7 subjects out of 9 show the mean values of 0.45, meaning that the voiceless responses are more frequent, while 2 answer the voiced sounds as 0.58 of mean values. It suggests that subjects tend to response the voiceless more in word-final position.

5.3.1 The cue-matched experiment.

The first experiment is conducted on how the native speakers perceive the original voiced or voiceless sounds correctly. In each environment, the stimuli consist of voiceless sounds such as '00', and '000' as well as the voiced sounds such as '11' and '111'. The result shows that the ratio of correctedness reaches 97% in the word-final released condition, and 84% in the word-final unreleased condition.

The results imply that native speakers have a little bit difficulty in identifying the voicing in case of removing the release. It confirms that some limited numbers of cues (e.g., release burst) play the meaningful role in identifying the voicing. In this experiment, it is safe to say that word-final stops with the release burst can lead to the perceptual identification more easily, confirming that the perceptual effect could be stronger in word-final released rather than the unreleased condition. Again, voicing identification prefers the quality of cues rather than the numbers of cues as well as environmental condition.

5.3.2 The perception types for the cues

The perception experiment is conducted for both tests of cue-match and cue-robustness. The cue-match test measures how the specific cues influence the voicing decision, in which the targeted cues are compared with the voicing responses. Along with the cue-match test, the cue robustness test measures the perceptual robustness for the specific cues through examining how the targeted cues lead to the voicing response, not masked by the neighboring cues. The test compares the target cues with the voicing response, manipulating the cues which compose a series of the different voiced or voiceless cues.

In the procedure, in case that the target cue leads the whole voicing response, the score for the target cue is counted. For instance, if the subject

responses the voiced sound for the manipulated signals of 'the voiced vowel + voiceless closure + voiced release', it adds the score for the voiceless closure because voiceless closure leads to the voiceless response.

In the word-final release context, the subjects' responses for the voiceless sounds receive more frequency than those for the voiced sounds: 57% (244) vs. 44% (188). Following table presents the match ratio.

Table 7. Cue match percentage in word-final released condition

Cues	voicing	frequency	response*		match ratio	
			voiceless	voiced	ratio	percentage
Pre-consonantal vowel	voiceless	frequency	177	39	41	76
		percentage	**41**	**9**		
	voice	frequency	67	149	35	
		percentage	16	35		
Stop closure	voiceless	frequency	120	96	28	49
		percentage	**28**	**22**		
	voice	frequency	124	92	21	
		percentage	**29**	**21**		
Release	voiceless	frequency	145	71	34	61
		percentage	**34**	**16**		
	voice	frequency	99	117	27	
		percentage	**23**	**27**		
Total		frequency	244	188		
		percentage	**57**	**44**		

*The test ranges from 25% as the bottom line to 50% as the maximum line per cue because it is divided by two types of the voicing. If the percentage for particular cue approaches to 25%, it means that it could be masked by other cues. On the contrary, it is very powerful cue if the cue approaches to 50%.

The results show that there are predominately more responses for voiceless stops (57%) than for voiced stops (44%). 76% of the responses are closely related with the cue of pre-consonantal vowels: 41 % of the voiceless responses and 35% of the voiced responses. However, the cue of the closure duration occupies only 25% (28% of the voiceless response and 21 % of the voiced response).

In the production experiment, the cue of release is little significant signal in the voicing ($p = .784$). On the contrary, in this perception test which provided 35 msec of release, it leads to 61% of the voicing responses. It means that the release is superior to the stop closure in perceiving the stop voicing. If we consider that the duration of release takes 35 msec equally for

both voiced and voiceless stops, subjects' perception is dependent on the spectral cues, not on the durational cues. Repp (1979) reports that the low amplitude of the release cue causes listeners to perceive voiced stops, while the high amplitude to the voiceless stops.

Following table is the result of cue-robustness test. In this context, 324 responses are analyzed for the cue-robustness test.

Table 8. The result of cue-robustness test in the word-final released condition

Cues	voiceless response	voiced response	Mean
Pre-consonantal vowel	91	65	78
Closure	28	22	25
Release	59	37	48

*unit: %

The result shows that the perception ratio of the voiceless sound shows clearly higher frequency than that of voiced stops, in case that the cues of closure or release insert in the manipulated signals. It means that voiceless perception is more dominant in the syllable-final condition. Thus, we can conclude that word-final neutralization is triggered by the perception, not by the articulatory production.

In the unreleased condition, the response of the voiceless sounds is more predominant: 54% of voiceless responses (116/216) and 46 % of the voiced response (100/216). Following table is the result of the cue match ratio.

Table 9. Cue match percentage in word-final unreleased condition

Cues	voicing	frequency	response*		match ratio	
			voice less	voiced	ratio	percentage
Pre-consonantal vowel	voiceless	frequency	89	19	41	79
		percentage	41	9		
	voice	frequency	27	81	38	
		percentage	13	38		
Closure	voiceless	frequency	64	44	30	56
		percentage	30	20		
	voice	frequency	52	56	26	
		percentage	24	26		
Total		frequency	116	100		
		percentage	54	46		

*The test ranges from 25% as the bottom line to 50% as the maximum line per cue because it is divided by two types of the voicing. If the percentage for particular cue approaches to 25%, it means that it could be masked by other cues. On the contrary, it is very powerful cue if the cue approaches to 50%.

In case that the cue of the pre-consonantal vowels is given the voiceless signal, the ratio reaches to 41% regardless of other cues voicing. Also when the pre-consonantal vowels are given the voiced signals, the ratios drop to 38%. Thus, it is safe to say that 79% of the responses emerge from the cue of the pre-consonantal vowels. However, the match ratio of the voicing responses for the closure duration is a little bit behind as 56%: 30% of the voiceless response and 26% of the voiced response. It means that the cue of the pre-consonantal vowels is superior to the other signals.

Table 10. The result of the cue-robustness test in the word-final unreleased condition

Cues	voiceless response	voiced response	Mean
Pre-consonantal vowel	81	65	73
Closure (voicing during the closure)	35	19	27

*Unit: %

In this word-final non-release context, 73% of the voicing responses emerge from the cue of pre-consonantal vowels, while 27% from the cue of stop closure. In case that release cue is removed, only the cue of the vowel duration takes charge of the voicing decision.

6. Discussion

6.1 Cue hierarchy

The failure of a one-to-one correspondence between the acoustic signals and the phoneme perception causes the problems in stop voicing categorization. For instance, the release cue proved as the significant signal in the perception level is reported as one of meaningless cues in the production level.

Naturally these mismatch relations trigger hierarchy of the phonetic properties. The phonetic cues realized in the release and approach stage are more prominent than those appeared in the closure duration. Some research suggests the cue-hierarchy in that only some limited cues obtain the prominent quality. Kang (2005) tests the masking effect of post-vocalic consonants, hypothesizing that the assimilation in the cluster consonants is decided by the acoustic information. That is, the enough acoustic information in the post-vowel consonants masks the place information in the pre-vocalic

consonants so that listeners tend to have dependency on the acoustic information of the post-vowel consonants in the sequences of VCCV syllable (consonantal cluster syllable). It means that the cues occurred in the release stage are proved to be the powerful signals. Finally, even though the closure duration is one of the prominent cues itself, it is easily masked when other cues occurred simultaneously.

The cue-hierarchy is closely tied with the human beings' perceptual system. Remez (2001) suggests that listeners tend to ignore, amplify, or categorize the sounds. It refers to sound distortion in both auditory and perception phases. By following the hypothesis, native English speakers tend to favor the durational cues of vowels or VOT, leaving the closure duration to the least effect. Thus, the cue-hierarchy could be arranged under the asymmetrical relationship between production and perception.

Table 11. The comparison between production and perception

Environment	cue types	production*	perception**
word-final release	main cues	vowel duration (.000)	vowel duration (78%) **release** (48%)
	subsidiary cues	closure (.056) **release** (.784)	closure (25%)
word-final unreleased	main cues	vowel duration (.008) **voicing during the closure** (.036)	vowel duration (73%)
	subsidiary cues		**voicing during the closure** (27%)

* significant values

** mean values of cue-robustness

The table implies the hypothesis of the cue-hierarchy in the perception level. Through the articulating stages in which various phonetic cues appear, some cues related with the approach and release stage are arranged again in the perceptual level. For instance, in the word-final release condition, the durational cue of the pre-consonantal vowels is more powerful than the closure. Thus, the model of the cue perception could be designed as follows:

Table 12. The models of the cue perception in each stage

	acoustic stage	perceptual stage	phonology stage
word-final released	[pre-consonantal vowel] [silence during closure] [voicing during closure]	[pre-consonantal vowel] [release]	less clear t/d distinction

Word-final unreleased	[pre-consonantal vowel] [voicing during closure]	[pre-consonantal vowel]	least clear t/d distinction

In the table, the phonological roles for the targeted cues are determined by the perception, not by the phonetics or acoustics. The cues resulted from both levels of phonetics and acoustics are filtered in the auditory phase. Then, some cues are selected fitted for the environment: pre-consonantal vowels of approach stage and release burst of the release stage. The reason that they have the prominent quality perceptually results from the characteristics of durational and spectral cues realized on the vowel duration, formant transition, release, and VOT (Wright 2001). The characteristics of the superior cues are as follows: (1) the superiority of the durational cues is associated with the spectral features, (2) the phonetic cues which have influence on perception could be different by following the contexts, (3) cues realized in the approach stage is perceptually salient because of the prosody effect.

6.2 The robustness of the phonetic cues in the approach stage

It is proved that the phonetic cues realized in the approach stage are very salient perceptually. Steriade (1995) suggests that the onset cues appeared in the initial part of the pre-consonantal vowels provides rich acoustic information rather than the offset cues appeared in the post-consonantal vowels in a serious of VOV (a composition of vowel, obstruent, and vowel). Furthermore, onset cues including VOT are superior to the cues occurred in the interior or offset cues. However, this study suggests that other cues could replace the perceptual strength of VOT in some conditions.

On the whole, the cue match ratio of the pre-consonantal vowel reaches the highest as 78%, compared with 48% of release/aspiration and 25% of the closure duration. Note that the roles of aperiodic cues such as release or aspiration drop sharply in this environment when they involve in the voicing decision. It seems that cue quality could be changed by following the contexts or environments. It supports the previous studies that vowel duration plays the meaningful role because spectral cues such as the fundamental frequency and the first formant frequency occurred in the pre-consonantal vowels produce the most salience (Kingston and Diehl 1994).

The reason that the durational cue of the pre-consonantal vowels is so prominent is closely related with the pitch values and shapes as a member of the suprasegmental features. Note that in this experiment the targeted stops emerge from pairs in the same environment beginning with '[kæ_]'. It seems that the fundamental frequency occurred in the pre-consonantal vowels affects perceiving the voicing. It is well-known that the stressed vowels raise the F0 which leads to the voiceless perception. The spectral cue of the first

formant frequency leads to the perception of the voicing and articulating places. For instance, lower F0 along with short VOT duration lead to the perception as the voiced consonants.

In the case of post-vocalic stops, however, the reduced amplitude (voiced) or silence (voiceless) that results from the stop closure provides recovery time for the auditory nerve. Therefore, if the post-vocalic stop is released, there would be a boost in production during the release. It is clear that onset placement yields significant nerve shock. Plauché (2001) reports that sounds transfer to basilar membrane through inner ears which trigger the auditory nerves fitted into the specific frequencies. That is, the auditory fibers respond sensibly on high density frequencies in onset of the amplitude. If the sound stimuli preserve for some time, the nerve response keeps falling down over the time passage.

The abrupt change for the spectral features has significant influence on the perception. Wright (2001) reports that speech sounds are transmitted via the outer and middle ear to the basilar membrane, exciting auditory nerves that are tuned to specific frequency bands. When a stimulus with an abrupt onset of amplitude is applied, certain auditory nerve fibers respond at a higher rate. If the stimulus has a more gradual onset of amplitude, the response rate of nerve decreases but remains steady. This property of auditory nerves suggests that the auditory system is highly sensitive to dynamic cues. Thus, the mechanism detects the abrupt change which includes the release, or voicing initiate. In spite of the vowel duration which holds spectral cues, it is sure that the pre-consonantal vowels located in the initial part of the syllable imply more prominent perception.

The results from the perceptual experiment illustrate the importance of considering environmental factors in phonological process. The experiment agrees with the study of Chang and colleagues (2001) who suggest an onset advantage that may be one of the factors in determining the environmental preference for onset over codas, and may be the foundation for the NO CODA constraint in OT. It is worth noting that the effect was not symmetric across stop voicing in some environments; while subjects' responses to stop voicing show clear advantage in the word-final released condition, recognitions in the unreleased condition imply less clear distinction under unreleased condition.

These sets of findings suggest that not all can be treated as equally. A release cue that is established in the word-final released condition implies the perceptual strength rather than it occurs in word-medial (VCV) syllables. This means that asymmetrical cue robustness should understand considering the interaction with environments or contexts.

6.3 Asymmetrical effect on sound change

This study again confirms the hypothesis that the relationship between perception and production is weak or indirect. If the acoustics imply direct

relationship with perception, the latter should perceive the acoustic signals perfectly and then connect them with phoneme perception. The presupposition is denied because phonemic perception is categorical, while the acoustic realization is presented as linear. The perceptual system evaluates the sound as a form of category since categorical recognition prefers or neglects some particular cues. Thus, the system distorts the outer sensory to pursue the categorical perception. In the end, the distorted relationship drawn from the psycholinguistic factors supposes the cue hierarchy. In reality, both experiments in this study prove the theory of Nearey (1997) and Remez (2001) who suggest the weak relationship between auditory and perception.

It seems that the asymmetrical effect has some influence on sound changes. The studies on sound changes insist that sound confusion triggers the alternation, seeking the optimal direction of sound change (e.g., Chang et al. 2001, Hume and Keith 2001). These studies tend to support the weak relation theory between phonology and phonetics as well as articulation and perception, since the perceptual confusion may trigger the sound change. Chang et al. (2001) suggest that the sound confusion in the direction of 'ki → ti' occurs, but the opposition as 'ti → ki' does not, because the direction is set from the strong perceptual cues to the weak cues. In the direction of 'ki → ti', the noise cues resided in velar stops are stronger perceptually than those in the alveolar stops. Thus, only some prominent cues lead to the sound change.

7. Conclusion

The results of both experiments clearly support the asymmetry between production and perception. Even though phonetic cues measured in the production exert the similar statistical significance, they could be reordered or rearranged in the perceptual levels. Analyzing the cue influence, the pre-consonantal vowel is proved to be the most prominent cue in word-final stop voicing. However, the release stage including release burst also holds the rich perceptual information in perceiving the voicing.

Even same cues have different qualities depending on the environments or the contexts. For instance, the vowel duration is proved to be more important in the unreleased rather than in the released condition. It points out the importance of environment rather than the phonetic cue itself when we discuss the cue quality. The hypothesis of the cue-hierarchy is set up as follows;

(2) Cue hierarchy of the stop voicing

Approach stage (pre-consonantal vowels) >> Release stage (release duration and burst >> Closure stage (stop closure, closure voicing)

Cues appeared in the release and approach stages are proved to be the

prominent perceptual cues rather than those in the closure stage. It means that amounts of the perceptual information are not equal among phonetic cues produced by the production experiment. Through both experiments, the study supports the weak or indirect relationship between production and perception. That is, asymmetry triggers the cue-hierarchy among phonetic cues which could be changed depending on the contexts or levels.

However, there are some noteworthy limitations of the study. This study focuses only on quantitative cues (e.g., duration), not on qualitative cues (e.g., spectral information). It means that cue weighting could be affected more by qualitative features (e.g., F1, F2). For example, release cue which is suggested as the important evidence to support the hypothesis on the asymmetry has been investigated only in durational information. Other acoustic information such as amplitude, spectral frequency, formant structure may affect the voicing perception. It remains for future works to study the qualitative features for the release cue.

REFERENCES

BOERSMA, PAUL and DAVID WEENINK. 2010. Praat (Version 5.1.25) [computer software]. http://www.praat.org. Amsterdam: Institute of Phonetic Sciences.

BORDEN, GLORIA, KATHERINE HARRIS and RAPHAEL LAWRENCE. 1994. *Speech Science Primer*. Sydney: Willimas and Wilkins.

BYRD, DANI. 1993. 54,000 American stops. *UCLA Working Papers in Phonetics* 83, 97-116.

CHANG, STEVE, MADELAINE PLAUCHÉ and JOHN OHALA. 2001. Markedness and consonant confusion asymmmetries. In Elizabeth Hume and Johnson Keith (eds.). *The Role of Speech Perception in Phonology*, 79-101. London: Academic Press.

FLEMMING, EDWARD. 1995. *Auditory Representation in Phonology.* PhD Dissertation. UCLA.

HAYWARD, KATRINA. 2000. *Experimental Phonetics.* London: Pearson Education.

HILLENBRAND, JAMES and ROBERT GAYVERT. 2004. *Open-Source Software for Experimental Design and Control*. Ms. [http://homepages.wmich.edu /~hillenbr]

HUME, ELIZABETH and JOHNSON KEITH. 2001. *The Role of Speech Perception in Phonology*. New York: Academic Press.

IVERSON, GREGORY and JOSEPH SALMONS. 1995. Aspiration and laryngeal representation in Germanic. *Phonology* 12, 369-396.

KANG, SEOKHAN. 2005. The experimental approach to the English voicing contrast by Licensing by Cue. Paper presented at the 2005 international conference of Korea Linguistics Society, 134-146. Seoul, Korea.

KENT, RAY and CHARLES READ. 2002. *Acoustic Analysis of Speech*.

Madison: Singular Thomson Learning Press.
KINGSTON, JOHN and RANDY DIEHL. 1994. Phonetic knowledge. *Language* 70, 419-454.
LIEBERMAN, ALVIN and IGNATIUS MATTINGLY. 1985. The Motor Theory of speech perception revised. *Cognition* 21, 1-36.
LIEBERMAN, ALVIN, PIERRE DELATTRE and FRANKLIN COOPER. 1958. Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech* 1, 153-167.
LISKER, LEIGH and ARTHUR ABRAMSON. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20, 384-422.
NEAREY, TERRANCE. 1997. Speech perception as pattern recognition. *Journal of Acoustical Society of America* 101, 3241-3253.
PLAUCHÉ, MADELAINE. 2001. *Acoustic Cues in the Directionality of Stop Consonant Confusions.* PhD Dissertation. University of California, Berkeley.
REMEZ, ROBERT. 2001. The interplay of phonology and perception considered from the perspective of perceptual organization. In Elizabeth Hume and Johnson Keith (eds.). *The Role of Speech Perception in Phonology.* London: Academic Press.
REPP, BRUNO. 1979. Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language and Speech* 22, 173-189.
SERNICLAES, WILLY and PIERRE BEJSTER. 1979. Cross-language differences in the perceptual use of voicing cues. In Harry van der Hulst and Vincent van Heuven (eds.). *Amsterdam Studies in the Theory and History of Linguistic Science* IV, 755-764. Amsterdam: Amsterdam-John Benjamin B.V.
STERIADE, DONCA. 1995. *Positional Neutralization*. Ms. UCLA.
STEVENS, KENNETH. 1999. *Acoustic Phonetics*. Cambridge: The MIT Press.
WILLIAMS, LEE. 1977. The voicing contrast in Spanish. *Journal of Phonetics* 5, 169-184.
WRIGHT, RICHARD. 2001. Perceptual cues in contrast maintenance. In Elizabeth Hume and Johnson Keith (eds.). *The Role of Speech Perception Phenomena in Phonology*, 251-277. New York: Academic press.

Seokhan Kang
General Education Institute
Konkuk University
268 Chungwon-daero, Chungju-si, Chungcheongbuk-do,
Korea 380-701
e-mail: kangs45@kku.ac.kr

received: March 11, 2014
revised: April 6, 2014
accepted: April 9, 2014