A Review of Current Theories Concerning Phonetic Invariance: With Some Implications from Clear Speech*

Seung-Jae Moon

1. Introduction

Phonetics has been trying to explain how we produce and perceive speech. One of the classic issues of phonetics is the invariance problem: the lack of one-to-one relationship between a linguistic unit and its corresponding physical properties (Perkell & Klatt 1986). Phonetic segments typically show extensive context-induced variability. This is a very basic problem to be resolved. Unfortunately, however, we do not seem to have a firm grasp on the issue. Thus, this paper aims at firstly reviewing some of the theories addressing the invariance issue. Depending on what theory we subscribe to, our research will take different approaches.

Recently phonetics seems to attract the attention of many related fields. It is the second goal of this paper to argue that, after reviewing the theories, we need to try to specify quantitatively the style-dependent transforms that speech patterns undergo instead of to try to find an "invariant" cue, by showing some recent findings about clear speech. It should be of value for the development of the related fields such as automatic speech recognition, and text-to-speech conversion.

2. Current Theories in Speech Production and Perception

2.1. Phonetic Gestures

One of the earliest theories is the Motor Theory of Speech Perception

*This study was supported by a grant from the Advanced Research Program of the Texas Board of Coordination, grant No. BNS-9011894 from the NSF, and a grant from Ajou University (1993).

(Liberman, Cooper, Harris, and MacNeilage 1962; Liberman, Cooper, Shankweiler and Studdert-Kennedy 1967), which was recently revised (Liberman and Mattingly 1985). Liberman and Mattingly (1985) argue that the objects of speech perception are not found at the level of the acoustic surface. The reason for this is the phenomenon of coarticulation. According to these authors, coarticulation serves the purpose of making it possible to produce speech more rapidly. However, the improved transmission rate has the significant drawback of making the signal more complex, or more 'encoded.'

They claim that "phonetic perception is perception of gesture" (p. 21). Phonetic gestures correspond to 'the speaker's intentions' which can be derived from the signal:

> ···the gestures do have characteristic invariant properties, ···, though these must be seen, not as peripheral movements, but as the more remote structures that control the movements. These structures correspond to the speaker's intentions. (p. 23)

The question arises; how do listeners manage to extract the invariant gestural information from the acoustic signal despite coarticulation? Liberman and Mattingly (1985) hypothesizes that it is retrieved by a specialized phonetic 'module.' According to the theory, the function of this module is:

> ···to make the conversion from acoustic signal to gesture automatically, and so to let listeners perceive phonetic structure without mediation by (or translation from) the auditory appearances that the sounds might, on purely psychoacoustic grounds, be expected to have. (p. 3)
>
> ···it is plain that speech somehow informs listeners about the phonetic intentions of the talker···these (phonetic) intentions are represented in a specific form in the talker's brain, and that there is a perceiving module specialized to lead the listener effortlessly to that representation. (p. 9)

Exactly how the module works, however, is not specified in the theory. It is assumed to be there, innately, for the purpose of processing speech.

Another theory according articulatory gestures a privileged status can be

found in Fowler (1986) who advocates adopting the so-called Direct Realism perspective on speech perception. She, too, assumes that the object of perception is the articulatory gesture: "In speech perception, the distal event considered locally is the articulating vocal tract." (p. 5)

Following the Gibsonian theory of direct perception, Fowler explicitly rejects the idea that cognitive factors play a role in perception:

> …perception must be direct and, in particular, unmediated by cognitive processes of inference or hypothesis testing, which introduce the possibility of error." (p. 4)

More recently, the theory of Articulatory Phonology was proposed by Browman and Goldstein (1990). Also in that framework gestures have the status of invariants.

However, the term "gesture" in their theory has a slightly different meaning compared with that of previous gestural theories. According to Articulatory Phonology, gestures do not correspond to features or segments.

A gesture is dynamically defined: it invokes several "tract variables," each of which, in turn, recruits one or more individual articulators. These gestures serve as basic units of phonology. Gestures can differ in their parameters such as constriction degree or constriction locations.

When these gestures come together into an integrated larger structure, they can overlap.[1]The overlap of gestures that differ in magnitude of constriction and in timing can give rise to many processes seemingly unrelated such as coarticulation, syllable deletion, assimilation etc..

2. 2. Acoustic and Auditory Invariance

Yet another viewpoint is that presented by Stevens and Blumstein (1978, 1981) and Blumstein and Stevens (1979, 1980). Searching for invariant acoustic properties in the signal they made the claim that invariance can indeed be found in the acoustic signal :

> …However, unlike the context-dependent theory, it is hypothesized

[1] This kind of gestural overlap combined with sluggish response of articulators was also envisioned in previous theorizing (e.g., Joos 1948 (the theory of overlapping innervation waves), Öhman 1966, 1967).

> that the properties of speech can be uniquely and invariantly specified from the acoustic signal itself (…) and that these properties are closely related to the distinctive features. (Stevens and Blumstein 1981, p. 2)

To properly decode this acoustic invariance is the task of "detectors" sensitive to those invariant cues :

> An hypothesis concerning the existence of these invariant properties implies that the auditory system is endowed with mechanisms that respond distinctively when a particular property is present in the acoustic stimulus. Property-detecting mechanisms of the type postulated here integrate a set of different acoustic attributes to yield an invariant or distinctive response in spite of the seeming variability of the individual attributes. (S & B 1981, p. 33)

More recently, Sussman (1989, 1991) and colleagues (Sussman, McCaffrey and Matthews 1991) presented the notion of locus equation and suggested that it captures in a significant way acoustic invariance.

Sussman was able to extract "extremely linear regression functions … characterized by distinct slopes and y-intercepts as a function of place of articulation" (p. 1309) from 20 subjects. The data allowed him to claim that relational invariance could be found in the acoustic signal. Sussman et al. also hypothesize a specialized neural recognition algorithm, similar to Stevens and Blumstein's (1981) detectors, for the cues provided by stop bursts and the parameters of the locus equations.

2. 3. Interaction between Speaker and Listener

However, there is a class of alternative theories offering a different point of view on invariance. Instead of trying to find phonetic invariance underlying the great variety of phonetic forms, they try to offer explanations for the variability observed.

This approach uses assumptions that were already expressed before the onset of this century. The idea that speech production is constrained both by listener's demands for clarity and the speaker's readiness to simplify articulations is very old in linguistics. We encounter this idea in early discussions of the phonetics of sound change (Passy 1890; Grammont 1933; Rousselot 1891; Martinet 1955; See also Weinreich, Labov, and Herzog

(1968) for a summary of various views including those of H. Paul (1880)). For example, Bloomfield (1941) already noted our tendency for utilizing ease of articulation :

> …It is safe to say that we speak as rapidly and with as little effort as possible, approaching always the limit where our interlocutors ask us to repeat our utterance, and that a great deal of sound-change is in some way connected with this factor. (Bloomfield 1941, p. 386)

On the topic of "ellipsis and explicitness," Jakobson and Halle (1968) also made an observation regarding speaker-listener interaction. According to them the explicitness of an utterance varies depending on the context.

Karlgren (1962) drew attention to the need for studying spontaneous speech :

> Phoneticians all over the world, to make life a little easier for themselves, have to a very large extent been working with utterances provoked in controlled experimental situations. (Karlgren 1962, p. 671)

In his study of speech rate, he claimed that the variability is not random; rather it is the result of a process that considers the situation at hand.

His account of phonetic variation induced by speech tempo was developed from ideas from Shannon and Weaver's information theory (cf. work by Nooteboom 1991 and Eefting 1991) :

> …Speech rate depends on the information content, among other things, and empty talk can be transmitted very rapidly indeed. (p. 674)

In his (1970) review paper, MacNeilage acknowledged the enormous variability present in speech production, and departed from "most theorists" at that time in viewing it as "an elegantly controlled variability of response to the demand for a relatively constant end" (p. 184).

He brought up the concept of "motor equivalence" from non-linguistic fields and tried to incorporate it into speech production models. He pointed to the serious gap between two groups of theorists with respect to their interpretation of the versatility of the motor system :

> To some neurophysiologists it even seems that with a few ad hoc hypothetical assumptions "spinal cord physiology has provided enough theoretical mechanisms to explain an elephant dancing the twist […]." We are thus faced with the paradoxical situation in which one group of

> theorists considers the motor system to have minimal versatility in the control of speech, and the other group encounters more versatility than they can see a use for! (p. 187)

He argued that "speech production is controlled, in part, by the specification of targets in an internalized space coordinate system" (p. 189-190). Furthermore, he proposed that this control mechanism has two components :

> …, it is tempting to suppose instead that the open-loop component emits a context-independent command for an articulator to reach a certain position, and that closed-loop control circuits constantly sample the mechanical state of the articulator and adjust the command accordingly. (p. 190, italics by author)

There is accordingly a research tradition in phonetics that is aimed at explaining the variability of speech in a systematic way. Recently it seems to have gained more momentum. For example, in his study of segmental reduction in German, Kohler (1990, 1991) suggests that we view reduction phenomena as follows :

> …What we need is the application of this new way of thinking to sound reduction phenomena at the sentence level in connected speech in order to provide a principled explanation for their occurrence and manifestation in synchronic sound systems. Questions to be answered are whether segmental reduction is rooted solely in the variability of speech production, governed by an *economy of effort principle*, and/or whether *perceptual constraints* intervene to filter out those articulatory processes from among the possible ones that then become established in a speech community. (Kohler 1990, p. 71, italics mine)

After examining several cases of reduction, he concludes that reductions should be viewed as a continuous process resulting not only from production constraints ('motor economy'), but also from perception factors.

Diehl and his colleagues (Diehl 1991; Diehl, Kluender, and Walsh 1990) have developed what they call the Auditory Enhancement Hypothesis to try to account for "(a) how listeners perceive phonetic structure and (b) why phonetic structure takes the form that it does" (Diehl et al. 1990, p. 261).

They are in disagreement with theorists who postulate speech-specific mechanisms. Using data from several experiments in which similar results

were obtained from speech and analogous non-speech stimuli, they argue that the observations can and should be explained in general auditory terms.

They argue that many speech phenomena including the gestural and acoustic covariation of consonants and vowels can be explained more consistently by appealing to general auditory mechanisms rather than to purely physical or physiological factors.

2.4. The Link Between Speaker-Listener Behavior and Sound Structure

Nooteboom and Eefting, focusing on the behavior of speakers, proposed so-called the principle of cooperative behaviour (Nooteboom 1991), or the adaptation principle (Nooteboom and Eefting 1991; Eefting 1991), which supposed that speakers adapted their pronunciation to the estimated needs of their audience.

The assumption that there is a link between the adaptive behavior of speakers and the structure of sound systems is implicit in the work of Liljencrants and Lindblom (1972). Exploring the concept of distinctiveness, Liljencrants and Lindblom (1972) conducted an experiment to simulate vowel systems and to test the hypothesis that 'a principle of maximal perceptual contrast' contributes toward the formation of natural vowel systems.

Their examination of maximal perceptual contrast was based on "the belief that vowels can serve as more efficient carriers of differences in meaning as they become more dissimilar, and the risk of confusing them decreases" (p. 855). Indeed, they obtained some encouraging results and suggested new directions for future research.

> ···It seems reasonable to suppose that a vowel system which has been optimized with respect to communicative efficiency consists of vowels that are not only 'easy to hear' but also 'easy to say'. Consequently a further improvement of the present theory might be obtained if we found a way of quantifying and incorporating 'ease of articulation'. , (Liljencrants and Lindblom 1972, p. 856)

The coupling between on-line phonetics and sound patterns is the focus of Ohala's research (1974, 1986, 1988). He is in agreement with several other researchers in believing that there is something more than just the signal involved in speech perception (1986). In his theory of sound change,

Ohala heavily emphasizes the role of listeners. When there is a 'breakdown' in the speech perception process due to listeners' failure to undo the various distorting effects of various contextual factors in speech signals, sound changes ('variations in pronunciation from one generation to the next' Ohala 1986, p. 388) occur.

However, he does not completely rule out the possibility that the speaker may play a role in shaping sound patterns.

In attempting to explaining sound change, Ohala strongly objects to the notion of 'ease of articulation' and also to taking a "teleological" approach.

Arguing from evidence on sound changes, Ohala concludes that speech perception requires top-down processing in addition to signal information. Therefore, he argues, we should not expect to find invariance in the signal.

Similar notions have been extensively explored by Lindblom (1983, 1988, 1990) who proposed the so-called H & H theory. The theory is derived from the belief that spoken languages share, with other biological processes, constraints that are "teleologically" governed. This formulation appears to be at variance with the stance of Ohala (cf. above).

Lindblom claims that speech perception is a matter of discrimination, thus **sufficient contrast or discriminability** is what is required for successful speech perception (cf. Diehl's auditory enhancement theory).

And this discrimination process is facilitated by signal-independent information :

> The process of discrimination is facilitated by processes not in the signal and whose contributions show short-term variations. Accordingly lexical access is assumed to be driven also by 'knowledge', that is by **signal-complementary processes.** (Lindblom 1990, p. 404)

However, the way listeners use this signal-independent information is not through a cognitive process such as "reconstruction" in Ohala's terms. Lindblom maintains more of a 'direct perception' view in this regard, relying on 'resonant kinematics,' the concept of 'internal representation' as envisioned by Shepard (1984).

H & H theory is also based on the assumption that speech production results from an interplay between production-oriented (that is, economy) and output-oriented (that is, plasticity) factors.

Based on the various observations on production and perception mentioned above, H & H theory claims that the speaker tacitly has the listener

in mind as a target for communication, and is able to adjust his performance according to the needs of the listener. Therefore "speakers are expected to vary their output along a coninuum of hyper- or hypospeech." (Lindblom 1990, p. 403)

Given these statements, we see that the theory takes a stance on the invariance issue which is similar to that of Kohler and Nooteboom. It "suggests that the lack of invariance ··· is a direct consequence of this adaptive organization" (Lindblom 1990, p. 403).

2.5. Summary

Despite their differences in theoretical layout, all the above theories have one thing in common. In one way or another, explicitly or implicitly they all address the invariance issue: the lack of one-to-one relationship between a linguistic unit and its corresponding physical properties (Perkell & Klatt 1986).

With respect to their treatment of the invariance issue we can categorize the theories summarized above into two categories. One group shares the assumption that phonetic invariance is a reality and that eventually it will be found in the signal. Finding it is just a matter of looking at the signal in 'the right way': Motor theory, Direct Realism and Articulatory Phonology assume that phonetic invariance is gestural whereas Stevens' Quantal Theory of Speech assumes it is acoustic/auditory.

The other group, discussed in section 2.3.and 2.4., approaches the issue by turning it around, so to speak, that is by trying to account for the variability.

How can we tell which one is more adequate? To shed some lights on the question, we will switch our focus briefly to the research of "clear speech."

3. Clear Speech

There have been many studies on clear speech (Chen 1980; Chen, Zue, Picheny, Durlack & Braida 1983; Picheny 1981; Picheny, Durlach & Braida 1986; Uchanski, Braida, Durlach & Reed 1986; Moon & Lindblom 1994). When "clear speech" is defined as a speaking style used to the instruction to speak as clearly as possible, Moon & Lindblom (1994) shows that clear speech has less degree of formant undershoot than citation-form

speech, using 5 American speakers speaking [w_l] words with front vowels. And it involves not only more vocal effort but also active reorganization of acoustic patterns and the underlying articulatory gestures. (Refer to the paper for the detailed discussion of the results.) After examining the intelligibility of clear speech, Moon, Lindblom, & Lame (1995) confirms the claim made in the above mentioned studies that clear speech is also more resistant to noise. Thus it has clear perceptual advantages.

Everyday informal experience suggests that "clear speech" is invoked by a speaker to meet certain communicative and situational demands. Apparently speakers are able to change and modify their speech according to the needs of their listeners. The results of the studies mentioned above indicate that speakers are quite capable of doing so in an experimental situation. They show an ability to adapt successfully as demands for explicit signal information vary.

4. Discussion

Then what can clear speech tell us about the nature of speech communication? Why do speakers bother to change their speaking styles in certain situations? And what do the theories explored in phonetics tell us about clear speech? Do they predict the existence of a "clear speech" phenomenon?

In Browman and Goldstein's articulatory phonology model, the basic mechanism is the timing and the amplitude of articulatory gestures. Would such a model explain clear speech? The results of Moon & Lindblom study (1994) strongly suggest the hypothesis that in clear speech gestures have to be reorganized. Browman and Goldstein have not yet addressed the task of modeling the kind of reorganization of phonetic gestures found in clear speech. If clear speech involves reorganization, that is, introducing more severe constraints on the achievement of targets that are acoustic or auditory, the answer to the question above appears to be, as far as I can tell, no. The articulatory phonology model does not seem to include the mechanism that would be needed for "listener-oriented control." At least not yet. Browman and Goldstein explicitly state that even in casual speech all the lexical phonological units are present and gestures do not change. Only their magnitude is decreased and temporal overlap may increase. However,

it is not uncommon to encounter cases where gestures are completely removed especially in casual speech (Kohler 1991).

The results of Sussman et al. (1991) seem very promising for signal-based invariance theories. However, the consistent slopes for each of the three places of stops could equally well be seen as satisfying a criterion of "sufficient contrast" rather than physical invariance. It is difficult at this stage to decide which interpretation is more plausible. Further study including various contexts and different speaking styles can shed light on the issue. The question is: Would the locus equations remain the same for the data gathered in different speaking styles? If locus equation parameters change with speaking styles, but maintain a contrastive distinction as a function of stop place, then a "sufficient contrast" position is supported.

The theories discussed in sections 2.1. and 2.2. including that of relational invariance have yet to explain the variability induced by style variations. Most of them have concentrated on the encoding and decoding of segmental strings and have been based on data acquired mostly under laboratory conditions by asking speakers to keep their vocal effort and tempo constant and by making recordings in a quiet environment. Although this is a common and widely accepted method in phonetics, the speaker's choice of different articulatory and stylistic tactics is ignored and thus uncontrolled. Moon & Lindblom (1994) tries to address this problem by explicitly directing attention to style-dependent phonetic variation. And the results obtained show systematic patterns for the two styles. It would seem that the regularity would be hard to account for by applying one of the signal-based invariance theories. Lindblom (1990) is very clear on this topic:

> Let us pause to try to envision a hypothetical situation in the future. Imagine that the invariance problem has been solved and that one of the signal-based theories is the winner…Given this state of affairs, what will the status of "intra-speaker phonetic variation" be? It would be true, would it not, that once a successful new theory emerges, the notorious variability of speech signals would by definition vanish. That variability would turn into epiphenomena arising from our incorrect 1989 way of looking at speech signals. Nonetheless, as soon as we looked at speech as we do today, all the extensive modifications (coarticulation, reduction etc) would still be there and would appear as

> systematic as they do now. What will successful signal-based theories of invariance have to say about that variability?
> Do we really want our theories to succeed in solving the invariance problem without having anything to say about the structure of intra-speaker phonetic variation? No, the systematicity of phonetic variation is very real and demands an explanation. (Lindblom 1990, p. 432)

In Ohala's view of sound change, great emphasis is put on the role of the listener. He claims that listeners normally adapt to the situation by factoring out some of the contextual signal distortions. They do so by a process of active reconstruction. If listeners are capable of this kind of reconstruction and complex top-down processing, and if speakers are aware that their listeners have this ability, why then is there any need for speakers to overarticulate and to produce clear speech? In other words, possessing the powerful ability to judge the signal in terms of its contextual information, should the listener, not be capable of retrieving the intended information from a decently transmitted signal? And would the speaker, knowing this, need to invoke overarticulation and compensatory action? Unless there is a limit to the adaptive abilities of listeners there does not seem to be any need for "listener-oriented" clear speech transformations.

By assuming that the origin of sound change is associated with purely accidental misperceptions by listeners, Ohala explicitly takes a non-teleological approach and advocates the idea that it is entirely "mechanical and non-purposeful." It would seem that Ohala's sound change mechanism is intrinsically non-directional. However, as observed by many linguists, sound systems often offer examples of changes that can be seen as facilitating either the listener's or the speaker's task. For example, Liljencrants and Lindblom (1972) argued with some success that a principle of perceptual contrast may be at work shaping vowel systems. Diehl et al. (1990) show that sound inventories prefer combinations of phonetic dimensions that enhance auditory distinctiveness. Phonological assimilations exemplify changes that bring about what often appears intuitively to be "articulatory simplifications." While Ohala's insistence of non-teleological accounts of sound change is well taken, the question remains: What mechanisms impose directionality on sound patterns? What makes them choose perceptual and/or articulatory "solutions"? Are perceptual and/or articulatory solutions

predicted by the Ohala misperception model? There appear to be many questions and few answers in this area. Perhaps "clear speech" data can shed some light on the details of phonological selections.

Both Ohala and Lindblom attribute a significant role to signal-independent information ("detailed phonetic knowledge" (Ohala) and "internal representation" (Lindblom)) in speech perception. The way the information is utilized in those schemes, however, is not the same as seen in the section 2.4. In Ohala's theory, the information is retrieved by way of top-down processing while in H & H, it is accessed more automatically in a parallel fashion. However, neither theory specifies how the mechanisms work in detail.

Diehl raises a question to explanations of speech perception based on attributing a role to this kind of "speech-specific tacit know edge" ("knowledge of phonetic and phonological regularities (as well as higher-order linguistic facts) acquired through experience with one's particular language" Diehl et al. 1990, p. 263). He does not completely dismiss the role of that kind of knowledge in speech perception. His point appears to be that we do not yet have enough information about that knowledge. Therefore we are initially better off pushing a purely auditory account to its limits.

I agree that a general auditory explanation, when it can be given, may be more powerful and better motivated than an ad hoc one based on speech-specific knowledge, and that it would have greater scientific value. However, as Diehl himself acknowledges, eventually that knowledge should be incorporated in a complete model. We have plenty of evidence that tacit knowledge plays an important role in perception. One example from Korean comes to mind immediately. Due to neutralization, the Korean phonemes /t, s, c, c^h, t^h, h/ (c: voiceless palatal affricate, ch: voiceless aspirated palatal affricate, t^h: voiceless aspirated alveolar stop) become [t] at the end of syllables. Consequently, five different words /nat/ 'grain,' /nas/ 'sickle,' /nac/ 'daytime,' /nac^h/ 'face,' and /nat^h/ 'each' are all pronounced identically as [nat] in isolation. When they are followed by a particle starting with a consonant, they are not distinguishable. However, a listener can easily perceive the intended word with the aid of the context. This example does not have any factor of reduction or any other of the complexities typical of speech signals in it. Needless to say the situation will be even more confusing once those kinds of processes also come into play. We favor the conclusion that, in the most general case, speech perception models can hardly do without

relying on "speech-specific tacit knowledge."

On the other hand, I fully agree that our present understanding of the representation and use of this tacit knowledge is highly incomplete. This is a limitation shared by theories such as the ones proposed by both Ohala and Lindblom. How is the knowledge represented? How is it used in the process, and to what extent? Is it a top-down process based on computation and inference as suggested by Ohala, or is it more of an automatic and direct process resonating with "internal representations" as proposed by the H & H theory? These questions have yet to be answered. However, the fact that we do not yet have enough information about this component cannot provide justification for rejecting the use of it in modeling speech perception. Rather what is needed is more rigorous research on the subject. A way of measuring the contribution of tacit knowledge in a quantitative manner has to be developed. Until a specific pattern emerges, the question ought be investigated thoroughly and the task deserves to be taken seriously.

However, as is now, the findings about clear speech clearly offer a challenge to those theories which support the presence of physical invariance. If speech perception works as an identification process (as Fowler suggested: "the special role of the perceptual system is not to create it, but only to select it." (Fowler 1986, p. 13)[2]), and if there is physical invariance in the signal, all that is required to achieve correct perception is selection of the invariants associated with the lexical unit intended by the speaker. If that is really the case, then why clear speech? More precisely, why does clear speech involve more than an improved S/N ratio? Why does it also exhibit systematic formant changes towards less context-dependence? Given a better signal-to-noise ratio, the speaker should not need to speak more clearly at all since the invariant attributes should be present in the louder signal. Why would he make an effort to speak clearer (in the sense of making formant changes)?

[2] The entire paragraph is as follows: How do listeners recover phonetic structure from such a signal? One thing is clear; the functional parsing of the acoustic signal for the perceiver is not one into acoustic segments. Does it follow that perceivers impose their own parsing on the signal? There must be a "no" answer to this question for an event theory devised from a direct-realist perspective to be viable. The perceived parsing must be in the signal: the special role of the perceptual system is not to create it, but only to select it.

The very existence of clear speech seems to suggest a very specific organization of the perceptual processing of speech. It can be taken to imply that speech perception is based on discrimination among patterns in the listener's lexicon and that it is aided by signal-independent knowledge. When there is enough knowledge, a more reduced signal will do the job as long as it contains just enough to trigger the percept formation. When the knowledge cannot help much, the signal itself has to do most of the work and hence it has to contain maximal discrimination. It has to be very clear. If speakers are able to estimate the informational needs, they can take advantage of the situation and adjust themselves accordingly. In this way speech behavior can be expected to be regulated by the constraints and the give-and-take of on-line speaker-listener interactions.

The systematic nature of the clear speech transformations seems to suggest that speakers were able to control the distinctiveness of the signal when needed to become more intelligible. In the future, explaining the existence and the properties of clear speech will have to begin by specifying the state of the perceptual processor.

Thus the results from clear speech research appear to provide strong support for theories based on sufficient discriminability since they explicitly emphasize the contribution from signal-independent information. However, as mentioned in the previous section, the notion of "sufficient discriminability" has to be elaborated a great deal and quantified in these theories. "Sufficient discriminability" should be formulated in a manner which can be investigated and tested more objectively.

5. Suggestion

So far we have seen basically two different approaches to the invariance problem. One approach tries to find "phonetic invariants" whether it be gestural or acoustic/auditory. The other approach tries to describe the variabilities found in the real life speech.

And with the data from clear speech, it has been argued that the latter approach is probably the right way to pursue. Instead of looking meticulously at lab speech data to find something which is supposed to be there but never found so far, we should try to understand the transformations taking place through many different speaking styles. To do so, we have to

gather abundance of data from various speaking styles. It is probably every researcher's experience to find a certain, nice regularity in lab speech only to lose it in other speaking styles such as conversational speech.

Researchers of phonetic science should organize their efforts to cover, hopefully, the entire spectrum of speaking styles. For that purpose, we have to establish a certain way of quantifying speaking styles. It is the current situation that many scientists are working on a speaking style which has the same name (for instance, 'reading style') but in reality is entirely different. One's reading style is not even close to others' reading style. Once we agree on this problem of defining speaking styles, we then should move on to the next stage of data gathering.

We do not seem to have a generally accepted Phonetically Balanced Words set of Korean let alone the sentence set. This is where our efforts should be put. A standard data set should be shared to provide data which can be interchanged and compared.

With these efforts, then we should challenge those notorious variabilities observed in natural speech. It is a goal of this paper to persuade researchers in the field of phonetics to try to specify this style-dependent transforms speech undergoes instead of to pursue the elusive, or perhaps never existed invariants.

References

Bloomfield, L. (1941) *Language,* Henry Holt and Company: New York.

Blumstein, S. and Stevens, K. (1979) 'Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants,' *J. Acoust. Soc. Am.* 66, 1001-1017.

Blumstein, S. and Stevens, K. (1980) 'Perceptual invariance and onset spectra for stop consonants in different vowel environments,' *J. Acoust. Soc. Am.* 67 (2), 648-662.

Browman, C., and Goldstein, L. (1990) 'Gestural specification using dynamically-defined articulatory structures,' *Journal of Phonetics* 18, 299-320.

Chen, F. R. (1980) 'Acoustic Characteristics and Intelligibility of Clear and Conversational Speech at the Segmental Level', Master's Thesis, MIT, Cambridge, Mass.

Chen, F. R., Zue, V. W., Picheny, M. A., Durlach, N. I., and Braida, L. D. (1983) 'Speaking Clearly: Acoustic Characteristics and Intelligibility of Stop Consonants,' 1-8 in *Working Papers II*, Speech Communication Group, MIT.

Diehl, R. (1991) 'Listener-oriented constraints on articulatory organization,' *Proceedings of the XIIth International Congress of Phonetic Sciences* Vol. 1, 92-96.

Diehl, R., Kluender, K., and Walsh, M. (1990) 'Some auditory bases of speech perception and production,' *Advances in Speech, Hearing and Language Processing*, Vol. 1, 243-267.

Eefting, W. (1991) *Timing in talking : Tempo variation in production and its role in perception,'* Doctoral Dissertation, The University of Utrecht.

Fowler, C. (1986) 'An event approach to the study of speech perception from a direct-realist perspective,' *Journal of Phonetics*, 14:1, 3-28.

Goldstein, L. (1991) 'Lip rounding as side contact,' *Proceedings of the XIIth International Congress of Phonetic Sciences* Vol. 1, 97-101.

Grammont, M. (1933) *Traité de phonétique*, Paris, Delagrave.

Jakobson, R., and Halle, M. (1968). 'Phonology in relation to phonetics,' in Malmberg (ed.) *Manual of Phonetics*, North-Holland/American Elsevier : Amsterdam/New York.

Karlgren, H. (1962) 'Speech rate and information theory,' in Sovijärvi and Aalto (eds.) *Proceedings of the Fourth International Congress of Phonetic Sciences*, Mouton & Co.: Hague.

Kohler, K. (1990) 'Segmental reduction in connected speech in German: phonological facts and phonetic explanations,' in Hardcastle and Marchal (eds.) *Speech production and speech modelling*, 69-92, Dordrecht/Boston/London: Kluwer Academid Publishers.

Kohler, K. (1991) 'The organization of speech production: clues from the study of reduction processes,' *Proceedings of the XIIth International Congress of Phonetic Sciences*, Vol. 1, 102-106.

Liberman, A., Cooper, F., Harris, K., and MacNeilage, P. (1962) 'A motor theory of speech perception,' *Proceedings of the Speech Communication Seminar*, Stockholm 1962, Stockholm: Royal Institute of Technology, 1963, D3.

Liberman, A., Cooper, F., Shankweiler, D., and Studdert-Kennedy, M. (1967) 'Perception of the speech code,' *Psychological Reviews*, 74, 431

-461.

Kohler, K. (1991) 'The organization of speech production: clues from the study of reduction processes,' *Proceedings of the XIIth International Congress of Phonetic Sciences,* Vol. 1, 102-106.

Liberman, A. and Mattingly, I. (1985) 'The motor theory of speech perception revised,' *Cognition,* 21, 1-36.

Liljencrants, J. and Lindblom, B. (1972) 'Numerical simulation of vowel quality systems: The role of perceptual contrast,' *Language* 48:4, 839-862.

Lindblom, B. (1983) 'Economy of speech gestures,' in MacNeilage, P. (ed.), *The Production of Speech,* Springer, New York/Heidelberg/Berlin, 217-245.

Lindblom, B. (1988) 'Phonetic invariance and the adaptive nature of speech,' in Elsendoorn and Bouma (eds.) *Working Models of Human Perception,* London: Academic Press.

Lindblom, B. (1990) 'Explaining phonetic variation: a sketch of the H & H theory,' in Hardcastle and Marchal (eds.) *Speech production and speech modelling,* 403-439, Dordrecht/Boston/London: Kluwer Academid Publishers.

MacNeilage, P. (1970) 'Motor control of serial ordering of speech,' *Psychological Review,* 77 (3), 182-196.

Martinet, A. (1955) *Écibinne des Changements Phonéiques,* Switzerland, Buchdruckerei Winterhur AG.

Moon, S-J. & Lindblom, B. (1994) 'Interaction between duration, context and speaking-style in the reduction of stressed vowels,' *J. Acoust. Soc. Am.* 96:1. pp. 40-55.

Moon, S-J., Lindblom, B. & Lame, J. (1995) 'A perceptual study of reduced vowels in clear and casual speech,' *Proceedings of the XIIIth International Congress of Phonetic Science,* Vol. 2, pp. 670-673.

Nooteboom, S. (1991) 'Perceptual goals of speech production,' *Proceedings of the XIIth International Congress of Phonetic Sciences* Vol. 1, 107-110.

Nooteboom, S. and Eefting, W. (1991) 'To what extent is speech production controlled by speech perception? Some questions and some experimental evidence,' To appear in: *Proceedings of the ATR Workshop on Speech Perception and Production,* Kyoto, Japan.

Ohala, J. (1974) 'Experimental historical phonology,' in Anderson and Jones (eds.) *Historical Linguistics* II, 353-389, Amsterdam: North Holland Publishing Company.

Ohala, J. (1986) 'Phonological evidence for top-down processing in speech perception,' in Perkell and Klatt (eds.) *Invariance and Variability in Speech Processes,* Lawrence Earlbaum Associates: Hillsdale, NJ.

Ohala, J. (1988) 'Discussion of Björn Lindblom's 'Phonetic invariance and the adaptive nature of speech',' in Elsendoorn and Bouma (eds.) *Working Models of Human Perception,* London: Academic Press.

Passy, P. (1890) *Etude sur les Changementsė Phonėiques,* Paris, Librairie Firmin-Didot.

Paul, H. (1880) *Prizipien der Sprachgeschichte,* Halle: Niemeyer. Fourth edition, 1909; Fifth, 1920.

Perkell, J. and Klatt, D. (1986) *Invariance and Variability of Speech Processes,* Hillsdale, NJ: LEA.

Picheny, M. A. (1981) *Speaking Clearly for the Hard of Hearing,* Doctoral Diss.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (1985) 'Speaking Clearly for the Hard of Hearing I: Intelligibility Differences Between Clear and Conversational Speech', *Journal of Speech and Hearing Research,* 28, 96-103.

Picheny, M. A., Durlach, N. I., and Braida, L. D. (1986) 'Speaking Clearly for the Hard of Hearing II: Acoustic Characteristics of Clear and Conversational Speech,' *JSHR,* 29(4), 434-446.

Rousselot, P. (1891) *Les Modifications phonėiques du langage ėtudiėes dan le patois d'une famille de Celle-Frouin (Charente),* Paris, Welter.

Stevens, K. and Blumstein, S. (1978) 'Invariant cues for place of articulation in stop consonants,' -*J. Acoust. Soc. Am.* 64 (5), 1358-1368.

Stevens, K. and Blumstein, S. (1981) 'The search for invariant acoustic correlates of phonetic features,' Eimas and Miller (eds.) *Perspectives on the Study of Speech,* 1-38, Erlbaum Assoc., New Jersey.

Sussman, H. M. (1989) 'Neural coding of relational invariance in speech: Human language analogs to the barn owl,' *Psychology Review* 96, 631-642.

Sussman, H. M. (1991) 'The representation of stop consonants in three dimensional acoustic space,' *Phonetica* (in press).

Sussman, H. M., McCaffrey, H. A., and Matthews, S. A. (1991) 'An investigation of locus equations as a source of relational invariance for stop place categorization,' *J. Acoust. Soc. Am.* 90 (3), 1309-1325.

Uchanski, R. M., Braida, L. D., Durlach, N. I., and Reed, C. M. (1986) 'The perception of temporally modified conversational and clear sentences by hearing-impaired listeners,' *J. Acoust. Soc. Am.* 80, Suppl. 1.

Weinreich, U., Labov, W. and Herzog, M. (1968) 'Empirical foundations for a theory of language change,' in Lehman and Malkiel (eds.) *Directions for Historical Linguistics,* University of Texas Press; Austin & London.

ABSTRACT

A Review of Current Theories Concerning Phonetic Invariance: With Some Implications from Clear Speech

Seung-Jae Moon

One of the classic issues of phonetics is the invariance problem: the lack of one-to-one relationship between a linguistic unit and its corresponding physical properties. Various current theories in speech production and perception concerning the invariance issue are reviewed.

Those theories can be categorized into two categories: The first group shares the assumption that phonetic invariance is a reality. The other group approaches the issue by trying to account for the variability.

Each theory is reevaluated under the lights of newly formed data from the area of "clear speech" research. And it is argued that the latter position should be taken for speech research and a way of conducting phonetic research in a systematic fashion is proposed.

Department of English Language and Literature
Ajou University
San 5 Woncheong-Dong, Kwonsun-Ku,
Suwon 441-744, Korea