DIRECTIONALITY OF VOICING AND ASPIRATION IN INITIAL POSITION

CHIN-WU KIM

Speech production involves a large number of muscles, the majority of which are capable of independent control. Krmpotić (1959) painstakingly measured the lengths and the diameters of a dozen representative nerve fibers (axons) which supply neural impulses to the muscles involved in speech production, and tried to establish the order of different degrees of latency in the propagation time of nerve impulses among these axons. Since how soon the neural signals reach their destinations depends mainly upon two factors: (a) the length of the axon (the longer the axon, the longer the travel time), (b) the conduction velocity which is proportional to the thickness of the axon (the thicker the axon, the faster the propagation), Krmpotić indicated these latency values in terms of what she called an INDEX, which is the length over the diameter of the axon $\left(I = \frac{l}{d}\right)$. See Figure 1. Numerical values in Figure 1 do not represent any actual time units, but merely indicate relative degrees of latency, implying that the greater the index value, the longer the time it takes for neural signals to reach the designated muscle, *ceteris paribus*. It is to be noted that the transit time of neural signals is different among various axons, and that those signals travelling down the recurrent laryngeal nerves that control the intrinsic laryngeal muscles have the greatest index value.

In view of Krmpotić's findings, there are two assumptions one can make.

(1) The nerve impulses of a speech act leave the cortex simultaneously so that those travelling the axons of higher index values will reach their destination progressively later than those travelling the axons with lower index values. In the absence of evidence to the contrary, this is a natural assumption. Galambos (1962:106), for example, showed that a series of protective actions, e.g., a crouching maneuver, that one makes upon hearing a loud explosive sound has a certain temporal order which is parallel to the distance of the location of the muscle to be activated from the brain, i.e., one first closes his eyes, he then draws his chin against his chest; next, he brings his elbows together to the sides of the body, and finally, he bends his knees.

(2) In learned behavior that must require simultaneous actions of different muscles which have different latency indices, the order of neuronal firings must be different

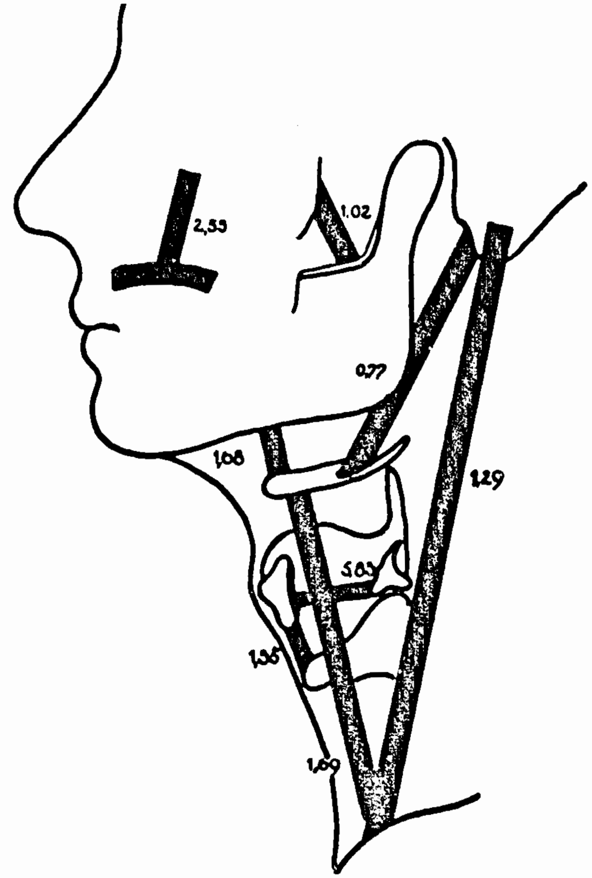

Fig. 1. Schematic representation of different indices of nerve impulse propagation time in axons leading to some muscles involved in speech production. Adapted from Krmpotić (1959:845, Figure 6). 0.77 — posterior digastric, 1.02 — trigeminal, 1.29 — accessory, 1.35 — vagus, 1.69 — hypoglossal, 2.55 — superior labial, 5.83 — recurrent laryngeal.

from that of the motor event. This is so because if the respective muscles are to be ready to contract simultaneously, that is, if a complex motor action is to come in time to produce an intended event, neural impulses to distant muscles must be fired earlier than those to the nearer ones to offset any time differential.

Which one of the two assumptions above applies to the act of speech? Lenneberg (1967:100-102) is of the opinion that the orders of neuronal firings are adjusted so as to achieve a temporal coincidence at the neuro-muscular juncture, and assumes that there exists in the nervous system an elaborate organization of interrelated neurons which, by some process of scanning, is capable of imposing certain types of synchronization upon a large number of widely spaced effector elements. He also attempts to explain the phenomenon of coarticulation in terms of different neuronal instructions to different segments being activated at a given time due to time differentials between motor and neuronal events. But if neuronal events are preadjusted to offset any temporal disorders in motor events, why should there be coarticulation at all? Lenneberg's schematic presentation is too abstract, and when one examines actual realization of phonic events, one even sees a glimpse of the first assumption operating. Even if it were the case that assumption (2) is operative in speech, it seems to me to be reasonable to assume that readjustment in the ordering of neuronal firings represents a 'marked' ordering which requires *a posteriori* learning by children, while assumption (1) represents a more natural and 'unmarked' ordering. We can further assume that since linguistic changes favor the unmarked over the marked ordering (cf. Kiparsky 1968), directionality of certain changes would be toward conforming with assumption (1) rather than with assumption (2).

Recall now that Krmpotić's finding showed that those neural impulses travelling down the recurrent laryngeal nerves would arrive at the intrinsic laryngeal muscles last. Since these muscles control the glottal closure for initiating vocal fold vibration, what the delay means is that in utterance-initial position, a voiced segment will become (partially) devoiced. It is reasonable to assume that a language learner has to learn how to prevent this devoicing, i.e., fire instructions to the glottis prior to those to the tongue or the lips. It is also reasonable to assume that the unmarked state will tend to prevail, giving rise to a progressively delayed closing of the vocal folds. This has a significant implication in directionality of initial voicing and aspiration.

Kim (1970) has defined aspiration as a function of the glottal opening at the time of the release of an oral closure of a stop. See Figures 2, 3. Thus, the less opening, the less aspirated; the more opening, the more aspirated. Now, if the neural signals to the intrinsic laryngeal muscles arrive later, the vocal folds would assume an open position that much longer, and to that extent the voicing onset for the following vowel would be delayed. This delay would make a voiced stop voiceless, and an unaspirated stop aspirated.

The above predictive assumption seems to be borne out well. In historical linguistic changes involving aspiration and voicing in word-initial position, the direction seems to be always [b] → [p] → [p^h]. Seldom is the reverse direction found. It has also been observed that children often speak with [h] or with aspiration, e.g., *hi* 'I', *hice cream*, etc. (H. Winitz, in a personal communication, pointed out to me that, according to the studies of Erwin, Lind, Trudy, etc., non-crying utterances of infants are made with the glottis in the normal wide-open position for expiration, thereby producing [h]-like sounds with great frequency, indicating that the skill required for glottal approximation has not been acquired.)

This serves as an example showing neurophysiological information contributing to formulation of an explanatory assumption about phonetic behavior, which in turn serves as a basis for a natural theory of phonology.

Department of Linguistics
University of Illinois
Urbana, Illinois

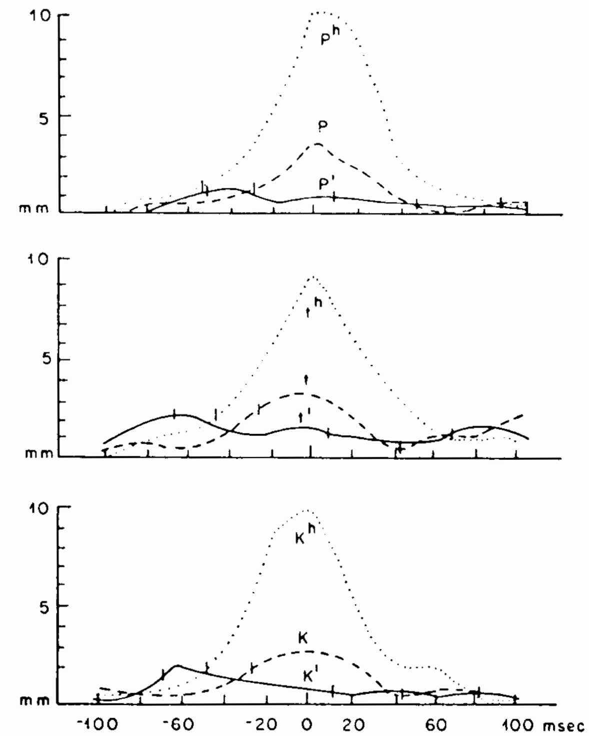

Fig. 2. Distance of glottal opening as a function of time in Korean stops. Traced from cineradiographic data (50 fps). For details, see Kim (1970). All curves are aligned with reference to the time of release as 0. Vertical bars before 0 indicate the time of oral closure, and those after the release indicate the time of voicing onset for the following vowel.

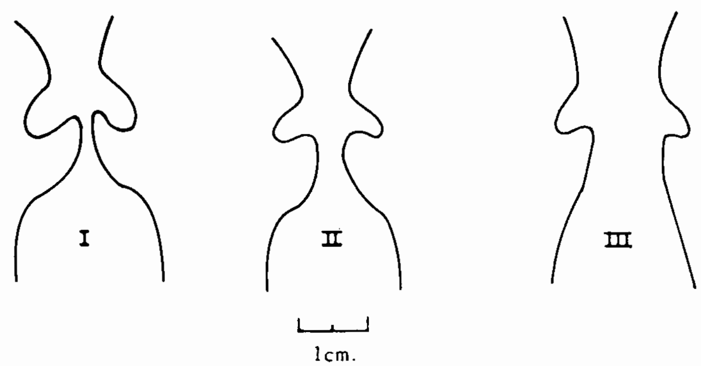

Fig. 3. Typical vocal fold shapes at the time of oral release in Korean stops: I — unaspirated stop, II — slightly aspirated stop, III — heavily aspirated stop. From Kim (1970:110).

REFERENCES

Galambos, R.
1962 *Nerves and Muscles* (New York, Doubleday).
Kim, C.-W.
1970 "A Theory of Aspiration", *Phonetica* 21:107-116.
Kiparsky, P.
1968 "Linguistic Universals and Linguistic Change", in *Universals in Linguistic Theory*, E. Bach and R.T. Harms eds. (New York, Holt, Rinehart and Winston), pp. 171-202.
Krmpotić, J.
1959 "Données anatomiques et histologiques relatives aux effecteurs laryngo-pharyngo-buccaux", *Revue de Laryngologie* 11:829-848.
Lenneberg, E.H.
1967 *Biological Foundations of Language* (New York, Wiley).

DISCUSSION

ABRAMSON (Storrs, Conn.)
You have cited diachronic evidence in support of your thesis. Is there any synchronic evidence? In a language with pre-voiced stops (voicing lead), we might expect to observe a tendency for voicing pulses to be missing from the occlusions quite often. Lisker and I have not seen this. In addition, your assumption of the simultaneous departure of neural signals for phonation and articulation may lean too heavily on the linguist's insistence upon neat segmentation of the speech signal.

KIM
The reason why the devoicing tendency does not happen synchronically in adult speech is because during the stage of language acquisition, children presumably learn to offset the time differential, whether it is due to the different arrival times of neural signals or to different response capabilities of peripheral organs. But I am assuming that a gradual accumulation of a tendency through many generations toward the more natural unadjusted state may eventually entail a phonemic shift. I am predicting that this shift is unidirectional for nonarbitrary reasons which I have stated in the paper.

GAGE (Washington, D.C.)
Why wouldn't the suggested mechanism lead equally to initial nasals becoming voiceless, a process which is not the prevalent direction historically?

KIM
A devoicing does not happen in the case of nasals, or laterals and vowels for that matter, probably because the voiceless nasals, voiceless laterals, or voiceless vowels are very rare and unnatural speech sounds.

LISKER (Philadelphia, Pa.)
If it is the case that a tendency to devoice voiced and to aspirate voiceless stops in initial position is the effect of a delay in the time of arrival of a neural command to the larynx relative to those activating the supraglottal mechanism, how would you explain the even more generally observed devoicing of final obstruents?

KIM
I think that the devoicing tendency in the final position is due to a different reason. In this case, the cause is probably anticipation, that is, anticipation of pause or the end of an utterance.

Since the resting state of the glottis is a wide open position for respiration, the vocal folds would tend to open toward the end of an utterance a little prematurely in anticipating of a pause, which will of course make the final segment devoiced. This phenomenon is analogous to the case of nasalization where a vowel is nasalized in front of a nasal consonant due to a premature dropping of velum in anticipation of the following nasal segment.

MACNEILAGE (Austin, Tex.)
Even if one makes rather generous assumptions about conduction velocity of neurons and the distance to be travelled to the larynx and the upper articulators — namely 50 meters per second in conduction velocity, and $\frac{1}{2}$ meter difference between the locations, the differences in time of arrival of impulses between the 2 sites is only 10 ms. Surely this is far too small to account for the differences in voice onset time between the different stop consonant classes which you are discussing.

Secondly, the nervous system could surely make up in power of motor innervation what it loses in time delay, in the initiation of our movement as opposed to another.

KIM
You're right. Ohala also reported that his reconstruction showed only 9 msec differential, not some 30 msec that Lenneberg computes. Perhaps an additional factor is the different response capabilities of different parts of the speech organ once the neural signals activated the respective muscles. The glottal movement seems to be much slower than that of the lips or the tongue. Various data (e.g., Daniloff, Kim, Lieberman) indicate that it takes as much as 100 msec for the vocal folds to close initially, while it takes only about 30 msec for the lips or the tongue tip to travel about the same distance.