

WE DO HEAR, FEEL, AND SEE OUR TONGUES AND LIPS

Rostislav Pazukhin

Pedagogical College, Czenstokhova, Poland

ABSTRACT

Speech researchers commonly disregard such modalities of oral speech as whisper, lip-reading, inner speech, etc., focusing on voiced speech exclusively. A survey of speech modalities shows, however, that this approach is not appropriate: as it turns out, the true bearers of linguistic information in all types of speech communication are *articulatory changes within the airway* (a cross-specific feature of all speech modalities). By contrast, the voice is usable in particular kinds of situations, where it serves only to reflect the articulatory gestures and make them audible to the addressee. The voice does not *produce* speech signals, but *reproduces* them.

PRELIMINARIES

1. My long observations of speakers of various languages in different situations have led me to the conclusion that our traditional view of speech production is inadequate. It is not „speech sounds” that are bearers of linguistic information in a speech utterance, but it is progressive changes within the airway that constitute the genuine speech signal. There are situations, where receivers perceive such *articulatory signals* directly (§§ 11,12). In most cases, however, articulatory manoeuvres and the resulting changes in the configuration of the speech channel remain hidden from our observation within speech cavities. In such cases, a kind of „echosounding” is applied:

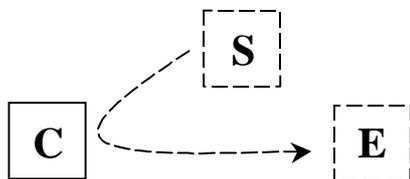


Figure 1.

Here, we have a „source” (*S*) which emanates a certain type of energy. The corresponding „echo” (*E*) reflects the invisible situation within our throat (*C*), see [19:29f.; 20:21-22].

2. The reality of Fig.1 may be demonstrated by cross-confronting the known *speech modalities*, such as whisper, lip-reading, inner speech, etc. All such modalities are avatars of a single general phenomenon, we call „oral speech”, and share certain basic codal and „technological” features. All speech modalities rely on the knowledge of a given language (English, Polish, etc.) and of the corresponding articulatory habits. Owing to this, the transition from one modality to another (say, *voiced speech* ↔ *whisper*) does not require from the speaker the study of a new language and of its specific articulations [19:30-31; 21]¹.

3. In what ensues, I present a short survey of the known speech modalities. It will be seen that only the scheme (Fig.1) is capable of describing and explaining each modality’s „inner mechanism” adequately. The kernel principle of such an explanation is:

(P1) *There may be no matter how many sources, sounding energies, and echoes of widely diversified physical nature. The efficiency of speech communication, however, depends not on the physical nature and shape of the echoes as such, but solely on the degree of precision, wherewith they reflect the inner configuration of the airway and its consequent modifications.*

SPEECH MODALITIES

4. **VOICED SPEECH** vividly illustrates the principle (P1). Voiced speech signal is not homogeneous in life, and undergoes innumerable modifications due to the age, sex of speakers, their psychic and physical condition, environmental factors, individual pronunciation [8:91; 22:i-ii]. As a consequence, researcher, for example, cannot find any postulated „invariant formant schemes” in concrete utterances [15:2102; 8:86]. The hearer can, however, identify these utterances, recovering modifications within the speaker’s speech channel from the chaotic acoustic images heard by him [21].

5. We also can recognise the spoken message in **SINGING**, **INSPIRATED SPEECH**, **SHOUTING**, **VENTRILOQUISM**, **MALE FALSETTO**, „**HELIUM**” **SPEECH**, despite their acoustic properties being extremely distorted and irregular [12; 14; 15; 23, etc.].

6. **WHISPER** is produced by air-friction whose noise-modulations mirror the changes within speech cavities [19: 33-34]. Without voicing one can thus replicate such “vocal” characteristics of speech as “vowels”, tonal accent, and even intonation [7; 9; 19:28-29].

7. There are **WHISTLED SPEECHES** (like the famous “Silbo Gomero”) which are “extraordinary realizations of the language spoken in the region where they occur”. Air eddies are produced by air stream on the lips, or on the fingers, and replicate feeble articulatory movements of the corresponding language [2].

8. “**FEMALE FALSETTO**” is not a voice, but a *whistle* generated by the air on the edges of the stiff vocal folds of a woman [25: 192, 195, 197]. It also reflects the movements of articulatory organs.

9. **ALARYNGEAL** forms of **SPEECH**. Patients with ablated larynxes do not usually lose their proficiency for speaking. The lack of natural vocal folds is partially made up by oesophageal sphincter, by remaining parts of the larynx, or by artificial reed-

appliances. This does not impede understanding the speech, despite the highly deviant characteristics of the sound [1].

10. LIP-SPEECH is based upon the *visual* echosounding: the light-rays emitted by the Sun, a bulb, or a candle, reflect the speaker's face movements. They, in turn, inform the receiver of the corresponding hidden articulations [6].

11. The TADOMA method is usually employed by deaf-blind persons, who have been taught to place their hands against the face of the speaker. In this manner, they can "read" articulatory movements directly. It is evident that no echosounding (cf. Fig.1) is needed here [5:401].

12. Nor are necessary any „echoes” in the INNER SPEECH, which is what we usually call the "tacit verbal thinking". It is argued that, here, afferent nervous impulses inform the „speaker” (or, rather, „thinker”) about the rudimentary movements of his own speech organs, corresponding to the „full articulations” of his own language [17:270-279]. It is, sometimes, also supposed that in such situations a direct interaction of efferent and afferent impulses may take place [10: 120-121].

DISCUSSION

13. Generally, speech students disregard the above described speech modalities, except for voiced speech. On rare occasions, however, they explain their preference for acoustic methodology by the wide application of voiced speech in everyday life and by a low practical importance of the remaining modalities [8:7; 25:212]. This conclusion is fallacious for at least two reasons.

First, one should not confuse two different planes of speech investigation. In the *explanatory* plane, the intrinsic nature of speech cannot be defined adequately, unless the properties of *all its known* modalities are taken into consideration. Such definition, according to the rules of classification [13], must be based on the properties common to all speech modalities (in this case: the articulation). By contrast, the voice is a property common to merely a chosen sub-class of speech modalities, being unfamiliar to the rest of them.

Second, in the *practical* plane, the domination of voiced speech should not be exaggerated. One should not forget the *inner speech* (§12), which functions as a feedback process controlling all kinds of speech communication including those which are soundless.

14. The above confrontation of speech modalities suggests a radical revision of our views of the mechanisms of *speech production*. It proves the inapplicability, in speech research, of the popular „Theory of acoustic filters”, which describes the process of speaking as producing conventional sound-signals with the bound acoustic properties [18: 5-7; 24, etc.]. Every speech modality may be, by contrast, described adequately by the "Echo-theory of speech" [20], which posits that speaking is the producing of conventional complex articulations. These are usually perceived by the addressee either proprioceptively, or by hearing, by sight, and by touch.²

NOTES

1. There were known attempts to attribute the inner homogeneity of speech modalities not to the shared articulatory nature, but to a postulated

"Precategorical Acoustic Storage". This imaginary psychic mechanism, allegedly common to all users of a given language, was proclaimed responsible for the "phonological" interpreting of non-acoustic forms of speech. All efforts to explain the functioning of the enigmatic PAS, however, engendered but scholastic and impracticable schemes [3; 4].

2. Also in the synthesized speech the problems of intelligibility and of naturalness cannot be solved, unless "echoes" of the imaginary articulations are imitated in the sound.

REFERENCES

- [1]. Ainsworth, W. and W. Singh. 1992. Perceptual comparison of neoglottal, oesophageal and normal speech. *Folia Phoniatrica*, 44, 292-307.
- [2]. Busnel, R. and A. Classe. 1976. *Whistled Languages*. Berlin: Springer.
- [3]. Campbell, Ruth. 1987: Common processes in immediate memory: Precategorical Acoustic Storage and some of its problems. In Allport, A. et al., *Language Perception and Production*. London: Academic Press.
- [4]. Crowder, R. 1983. The purity of auditory memory. *Philosophical Transactions of the Royal Society of London*, B 302, 251-265.
- [5]. Crystal, D. 1987: Tadoma method. In Crystal D. (ed.). *The Cambridge Encyclopedia of Language*. Cambridge UK: University Press.
- [6]. Ewing, Irene. 1944. *Lipreading and Hearing Aids*. Manchester: University Press.
- [7]. Fonagy, J. 1969. Accent et intonation dans la parole chuchotée. *Phonetica*, 20, 177-192.
- [8]. Kent, R. and Ch. Read. 1992. *The Acoustic Analysis of Speech*. San Diego CA: Whurr.
- [9]. Kloster Jensen, M. 1958. Recognition of word tones in whispered speech. *Word*, 14, 187-196.
- [10]. Lashley, K. 1951. The problem of serial order in behavior. In Jeffer, L. (ed.), *Cerebral Mechanisms in Behavior*. New York: Wiley.
- [11]. Mendel, L. et al. 1995. Speech intelligibility assessment in a helium environment I. *JASA*, 97, 628-636.
- [12]. Mendel, L. et al. 1998. II. *JASA*, 104, 1609-1615.
- [13]. Mill, J.S. 1865: *System of Logic, Ratiocinative and Inductive*, Vols. I-II. London: Longmans-Green.
- [14]. Miller, R. (1986) · *The Structure of Singing*. New York: Schirmer.
- [15]. Neary T. 1989. Static, dynamic, and relational properties in vowel perception. *JASA*, 85, 2088-2113.
- [16]. Orlikoff, R., R. Baken, D. Kraus. 1997. Acoustic and physiological characteristics of inspiratory phonation. *JASA*, 102, 1838-1845.
- [17]. Ostwald, (1973). *The Semiotics of Human Sound*. The Hague: Mouton.
- [18]. Owens, F. 1993. *Signal Processing of Speech*. London: MacMillan.
- [19]. Pazukhin, R. 1993. *Oral Speech: Controversies and Perspectives*. Lodz: Author's Edition.
- [20]. Pazukhin, R. 1996. A linguistic-acoustic-computational analysis of spoken communication: Illusion and reality. In Pazukhin, R. *Topics in Semiotics and Linguistics*. Czenstokhova: WSP.
- [21]. Pazukhin, R. The recent crisis in speech acoustics and its possible implications for linguistic theory. In Caron, B. (ed.). *Actes du 16e Congrès International des Linguistes (1997 Paris)*, CD ROM, Oxford: Elsevier Sciences, (in press).
- [22]. Perkell J. and Klatt, D. (eds.). 1986. *Invariance and Variability in Speech Processes*. Hillsdale NJ: Erlbaum.
- [23]. Rubin, H. and Ch. Hirt. 1960. The falsetto: a high speed cinematographic study. *Laryngoscope*, 70, 1305-1324.
- [24]. Stevens, K. 1997. Models of speech production. In Crocher, M. (ed.). *Encyclopedia of Acoustics*, vol.IV. New York: Wiley.
- [25]. Zemlin, W. 1968. *Speech and Hearing Science: Anatomy and Physiology*. Englewood Cliffs NJ: Prentice Hall.