THE PERCEPTION AND COMMON MISPERCEPTION OF INFANT PRE-SPEECH

H. M. TRUBY*

INTRODUCTION

In an attempt at harmony with the announced central theme of this Congress, the following observations are here offered for consideration. Acoustic signal, after all, is tangible. Its perception is most demonstrably *not!* The specification, acoustic speech signal, *can* be held to the tangible, but is not usually so restricted, since most evaluation is based on phonemic and higher level linguistic principles than the uninvolved, purely physical (to include biophysical) analysis entails. *Phonological* implications, once admitted, alter the course of objectivity... quite beyond recognition, it can fairly be added. Were this not so, the dichotomy implied by the two rubrics, *phonetics* and *phonetic sciences* might never have come to be entertained. Speech is the concern of phonetics—the speech signal is one of the concerns of the phonetic sciences. It is impractical—if not impracticable—for the phonetician to avoid dialectology or phonemics by whatever names. Various departments of the phonetic sciences are, interrelatedly, directed at various physical—to comprehend biophysical—areas. The measurements which define these departmental interests should depend in no way upon perceptual proclivities, and one of the most obvious and accessible and expounded of these measures is the acoustic. It is troublesome—but of little consequence—that the disciplines of phonetics, of the phonetic sciences, and of phonology are rarely treated in mutual exclusion. It is, however, of considerable consequence that *physi*cal inferences are quite often drawn from *psycho*physical exercises. This paper is directed at illustrating the consequence of such misappropriation.

DISCUSSION

The developmental sector, "infant pre-speech", is directly the concern of phonetics. More has been miswritten about developmental speech than a dozen papers of this scope could hope to offset. By selecting the *pre*-speech sector, linguistic criteria are by definition irrelevant. By admitting perception—and misperception—the full

* Communication Research Institute, Coconut Grove, Florida. (Now, Professor of Pediatrics and Linguistics, University of Miami, Coral Gables, Florida.)

complement of psychophysical vectors and parameters may be treated. Let us now consider in what ways these conflicts breed invalidity.

When acoustic signal is mentioned, sound is explicit. Where there is sound, there is motion to which that sound relates... from which it derives physically.[1] In the human, where there is conventional speech-sound motion, there is neuromuscular transmission, purely neural pregeneration, and something that may as well be referred to as conceptual *pro*generation. All of this biophysical activity is as deserving of the name *speech* as is the resultant acoustic activity, and the same holds for the progenerating conceptualizing process.[2] Yet, insistence on such definition is clearly moot. [The roles of feedback or of the listener or receiver are purposefully avoided here.] Our present objective is the derivative relationship of the acoustic signal to its generatory motion.

Sound is enigmatically elusive insofar as memory registration is concerend.[3] There simply is no way to describe most sound continua. Speech sound is, to a degree, ont of the rare exceptions, but even with speech sound, not only is the pattern-conditioning literally indescribable, but there is much of the speech-sound signal which is neither recoverable nor registerable. Anyone who has ever attempted to re-create or just plain remember the voice of an acquaintance or a family member or whomevefinds it impossible to proceed. Which criteria can be used? True, certain timing and emphasis and voice quality features can be more or less indicated, but on the whole, a faithful physical recounting is not to be had. Yet it cannot be denied thar the recognition aspect, for example, is clearly no less real than the identificatione with-certain-phonemes aspect. How does such a situation come to be? The answer is in the pattern-matching criteria directly, society having provided the incentive of communication to speech-element identification generally, while the phenomenon of speaker-recognition has remained on the whole a more individual or small intimate group consideration, public figures perhaps to the contrary. Yet one sound aspect is as physically real as the other—voices can certainly be identified by those familiar with the details of their acoustic manifestation, and if they are permanently recorded in some audible form, they are permanently available for re-identification. Otherwise, there is *no* physical medium form in which the details implicating the inherent acoustic features may be described differentially—there is no sound-symbol system keyed to voice quality minutiae. Alphabets are basically speech-sound-symbol systems, but no one would for a moment attempt to transcribe voice-quality differences with conventional alphabet symbols. Nor would the phonetician have any

[1] Cf. Truby and Lind, "Cry Sounds of the Newborn Infant", *Acta Pædiatrica* 163, 1965, p. 9, and see Footnote 3 hereswith.

[2] Truby, "A Definition of Speech-sound Analysis, "Speech Synthesis", and Speech", André Martinet Festschrift, LINGUISTIC STUDIES, Clowes, N. Y. — London, 1968.

[3] Truby, Bosma, Lind, *Newborn Infant Cry, Acta Pædiatrica Scandinavica* 163, 1965, p. 17. (first published as Communication Research Institute Scientific Report No. 0260, 1969).

more success with any so-called phonetic alphabet. Yet, with complete disregard for the irrelevance of it, more than a few students of speech have transcribed—in terms of speech-sound symbolism—a wide variety of events to which their only analysis access was auditory. The case in point is the catalyst for this report.

Physiological phonetics has traditionally made an issue of the fact that speech is an overlaid function of respiration, deglutition, even mastication, and the like. However, the focus has quite understandably always been upon articulation, with some spillover into the aspects of source generation, intonation, emphasis, and timing, for instance. Whenever subphonemic details are marked, they relate in turn to articulatory features primarily and associated physiological aspects secondarily, but all with the *basis* in mature physiology and mature performance. Consider, if you will, the all-important area of newborn-infant cry-sound analysis. Various combinations of direct observation, X-ray motion-picture photography, sound-spectrographic records, and the like reveal that the infant uses a wide variety of compensatory articulatory and sound-generation mechanisms to accomplish what seem to be "linguistically same" vocal productions. The representation of these performances using conventional phonetic transcription (the basis of which is, as stated, in adult physiology and anatomy) is demonstrably *in*appropriate to the task of *pre*-speechsound evaluation. From the purely physical point-of-view, cry-sound sequences have "vowels" and "consonants", and nasalization, and voicelessness, and they differ from each other in their spectral envelopes, and durations, and intonational effects as well, *but*, cry sound should certainly not be analyzed as though it were speech. In fact, much damage has been done by observers who have assumed that when they—the observers—perceive the presence of a particular sound image in an infant emission, they should indicate in their transcription and in their evaluation the articulatory mechanism conventionally associated with that image. Nothing could be further from the facts, which are, essentially, that differing articulations will not produce the phonetically same oscillogram or sound spectrogram, no matter how phonemically alike these sounds may appear to be! Phonemics simply has no place in the analysis of the non-speech signal, and extreme care is invited wherever compensatory mechanisms are apt to be general (as in pathology) or wherever analogs are employed (as in so-called "speech synthesis").

The archives of this researcher contain thousands of recorded cry-sound sequences, many of which were made with correlated cineradiographic monitoring. Sound spectrograms of the tape-recorded events, as the recorded events themselves, carry acoustic cues which, for the native English speaker, would seem to correspond to conventional speech-sound sequences. For instance, hundreds of sequences would appear to be describable as [mɛ̃ː], (see A of figure) yet, inspection of the cineradiographic print of any of these sound productions would verify that the infant's lips never even approximated bilabial occlusion and, in fact, that his mouth was rather wide open, the tongue suspended under tension and concave in more or less low central position, and that all relevant articulation—that is, all articulation resulting in the

perceived bilabial nasal continuant—took place in the mesopharynx and consisted of very evident approximation of mesopharyngeal wall to tongue-root! Practically all egressive, infant cry sound is characterized by the tongue position described, insofar as the most vowellike portion is concerned, and practically all activity which could be construed as articulation takes place in the mesopharyngeal area as illustrated above. Thus (see B of figure), there is no alveolarity in articulatory evidence during a cry-sound sequence heard as [næ̃::] ("dentality" being ruled out in the absence of teeth!), and so on throughout any inventory correspondences imagined by a listener (see, for example, C of figure). Medical doctors unacquainted with basic phonetic principles can be excused for their misevaluations (though not forgiven for not seeking out professional counsel), but it seems a shame that reputable phoneticians-by-portfolio must also be included among the culpable. The, I hope unforgettable, lesson-to-be-learned is: "Do not assume an articulation or other phonetic feature on the basis of acoustic evidence alone!" The acoustic facts may well be objective and valid and reliable, but phonetic identifications by observers are subconsciously and almost ineradicably based on habitual adult articulatory performance and have little or no reference to sound-producing mechanisms per se.

All infant vocalization leads, in some sense, to the ultimate stabilization, essentially, of speech-performance patterns, and attempts to account for the unfolding details of these developmental pathways have proved intriguing to many professionals and non-professionals alike and will continue to prove so, but there is good reason to proceed cautiously as suggested above. At certain later stages of what may be termed "pre-speech" performance, there is relevance to phonetic evaluation in terms of adult linguistic-signal processing, but at earlier stages, such evaluations are generally, if not usually, highly irrelevant. It is not likely that technological instrumentation will be developed to account for the infinite variety of phonetic interpretations, and all premature evaluations (by investigators, and in the senses indicated) are by definition suspect.

That which has been offered as regards newborn-infant sound production could, with impunity, be extended, as every linguist amoung us will corroborate, to any and all *inter*-language considerations, and the International Phonetic Association made itself irrefutably clear on this point right from its beginnings. The significance of the IPA assertion that visual symbols are only by convention referable to particular vocal events has been essentially lost in the last important area of misappropriation which this report shall treat, and that is the area which has come to be known as "speech synthesis". Since what has to be said in this regard is at once obvious in the light of the material discussed above, it must suffice to say that the *product*—the signal output—of the synthesizer might more aptly be termed *pseudo*speech, and the implications will be at once evident to every member of this Congress. Or certainly *should* be! Machines simply do not have teeth nor lips nor tongues nor the like, and the sort of transcription which applies to their output is a bird of a different breed. Admittedly the output is audible, and insofar as the genius of the programmer is

concerned, the machine program is related to human programs. At that point the analogy ends.

In order to approach the conclusion of this paper with a highly relevant and serious consideration, may all of its consumers be reminded that the success or failure of treatment in pathology or the validity of research rest with the phonetic diagnosis. An infant with loss of neuromuscular coherence in the tongue-blade area whose output is evaluated by the phonetician in terms of apico-alveolar performance, has little chance of being exposed to relevant therapy until the phonetician's misdiagnosis is discovered. It is not enough to have "a highly trained ear", nor, in the case of sound spectrograms, a "skilled eye". Nothing can take the place of an immaculately plenary consideration of all the possibilities. It is to be hoped that the suggestions offered here will assist the phonetician in capitalizing on the full store of information his special training has equipped him to uncover.

CONCLUSIONS

Conventional phonetic transcription,[4] based as it is on mature-speaker anatomical, articulatory, and associated speech-feature particulars, is inappropriate to the task of describing *immature* (in whatever sense) sound performance,[5] which though emanating, as speech, from the upper respiratory tract, may be seen to be more remotely constituted *from* speech than is immediately apparent. This does not mean to infer that there is not much of great interest to be considered as regards the developing program of employment of varying (with maturation) physiological mechanisms from speech-sound to speech-sound and from individual to individual. However, as suggested in the opening thoughts of this paper, the acoustic signal is, in a sense, tangible—the perception of that signal, being a psychophysical operation, is at once arbitrary, illusionary, and based, at best, on an unstable inventory of phonological bias. It is thus best to proceed with caution where the objective is physical description. *Communication* is one thing—the measurement of *modi* and *media* is quite another.

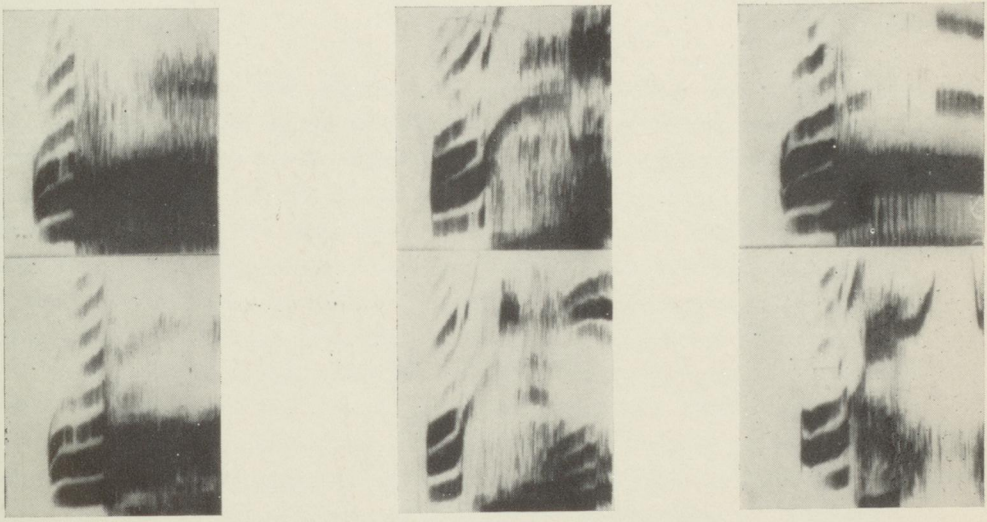
---

[4] Cf. Truby, "Pleniphonetic Transcription in Phonetic Analysis", *Preprints* (Harvard-MIT, 1962) and *Proceedings* (Mouton, 1964) *of IX International Congress of Linguists* and the "phonette" in *Acoustico-Cineradiographic Analysis Considerations, Acta Radiologica* 182, Stockholm, 1959.

[5] Truby, "Cry Sounds of the Newborn Infant" (op. cit.), pp. 17—18, 57. "Language and Dolphins", Xth International Congress of Linguists, Bucharest, 1967.

Slama-Cazacu:

Les deux communications (Mme Sedláčková et M. Truby) touchent un problème qui commence à être exploré avec des procédés objectifs. Des recherches de ce genre sont utiles non seulement pour le problème du développement du langage, mais aussi pour la phonétique générale. Elles m'ont intéressée aussi d'un point de vue personnel: nous avons initié une recherche, en collaboration internationale, concernant la formation du système phonématique chez l'enfant et des recherches de détail, si minutieuses, comme celle de Mme Sedláčková, nous seront très utiles. Je voudrais souligner aussi la nécessité de réaliser — par une coopération — un système international de transcription phonétique adéquat au pré-langage enfantin.



A B C

A — Sound spectrograms of newborn-infant cry-sound initiations resembling [*mæ̃ː*], but see related text!

B — Sound spectrograms of newborn-infant cry-sound initiations resembling [*næ̃ː*], but see related text!

C — Sound spectrograms of newborn-infant cry-sound initiations resembling [*ŋæ̃ː*], similarly to above cited circumstances.