COOPERATIVE AND INTERACTIVE MODELS OF HEARING
AN INTRODUCTION

MANFRED R. SCHROEDER

Drittes Physikalisches Institut, Universität Göttingen
Bürgerstr. 42-44, D-3400 Göttingen, West Germany, and
AT&T Bell Laboratories, Murray Hill, New Jersey 07974, USA

Since the days of Ohm, Seebeck and Helmholtz, we have gained great insight into the functioning of the peripheral auditory system of man (and many other species). We finally begin to understand even Tartini's (1740) "terzi suoni" as a nonlinear byproduct of active mechanisms in the inner ear.[1]

Highly sensitive *physical* measurements led to the discovery of "Kemp echoes" and oto-acoustic emissions. Combined with intracellular (hair cell) recordings and other *physiological* triumphs, the resulting insights have taught us what we have suspected all along: the mechanical resonators in the inner ear (von Békésy's traveling waves as seen in post mortem specimens) are but *pre*filters for delicate active processes to build on. Only thus can we understand the astounding sensitivity of our ears (close to the detection of random molecular motion) and their great frequency resolution (almost--but not quite--violating Heisenberg's Uncertainty Principle).

Psychoacoustic research has likewise made great strides in decades past and has contributed enormouly to our understanding of hearing. *Mathematical models* of auditory functions have had another heyday--and some models have even clarified our thinking about the ear.

While most stimuli used in hearing research have been of the nonspeech kind--the familiar tones, clicks and hisses in a veritable carnival of post-, pre-, and simultaneous masking paradigms, pulsation thresholds, loudness estimates etc.--a considerably corpus of knowledge using speech or *speech-like* stimuli has also been built up.

1. Not surprisingly, these "amplifiers" involve bio-*molecular* processes--what *else* could go nonlinear at basilar membrane "excursions" smaller than 1 Angstrom, the diameter of the hydrogen atom.

These experiments have taught us much about subtle interactions between the time and frequency domains and much else that is important for speech perception.

However, there remains a large white area of unexplored territory, located somewhere between these two continents explored by past research using speech- and nonspeech-like stimuli. Numerous experiments beckon us to enter this noman's land to discover properties of the auditory system that are *pre*-speech and yet transcend previous nonspeech work.

Why is it that a vibrato ("frequency modulation") imparted coherently to the individual components of a tone complex will tend to fuse these tones into a perceptual whole? Indeed, what *are* the physical parameters in complex stimuli that will lead to perceptual fusion on the one hand and, on the other hand, permit the separation of different "voices" (including those of musical instruments and other nonspeech signals).

Why is masking of a tone pulse influenced by the coherence ("phase stability") of the masker *hundreds* of milliseconds before or after the maskee [1]? What kind of temporal integration is at work here?

Considerable integrative capabilities of the human ear for nonspeech stimuli have also been observed in the *frequency* domain, both regarding gross spectral features ("formants") and spectral fine structure ("pitch). A recent model [2] addresses the latter problem, but much remains to be explored on spectral shape integration and timbre perception.

In this context, we might consider also the subnormal speech of hearing-impaired speakers as another kind of prespeech. Certainly, the results of research on recognition of such "speech" are confusing, to say the least, in terms of present psychoacoustic understanding [3].

The perceptual correlates of phonetic features and their *relational* properties [4] are another important subject that needs to be explored further in the arena of cooperative models of hearing.

Another area of concern is the interface between acoustic-phonetic analysis and lexical processes. W. D. Marslen-Wilson and U. H. Frauenfelder, in their contribution to this Symposium [5] argue strongly for integrating the methodologies of acoustic-speech research and psycholinguistics.

One of the strongholds of cooperative auditory processing is the human binaural system and we welcome a contribution with exciting new results to this important topic [6]

It is hoped that the present Symposium, in addition to auditing exciting new results from its participants, will provide a forum for focusing on those cooperative and interactive aspects of hearing that have been hitherto largely ignored. Wish that Tallinn will provide a strong stimulus in our search for new and revealing auditory stimuli!

REFERENCES

[1] Work by S. Mehrgardt, cited by M. R. Schroeder, "Speech and Hearing: Some Important Interaction," in M. P. R. Van den Broeke and A. Cohen (eds.), Proc. 10th International Congress of Phon. Sci. pp. 41-52 (1983).

[2] A. J. M. Houtsma and J. G. Beerends: "An Optimum Pitch Processing Model for Simultaneous Complex Tones," Symp. on Cooperative and Interactive Models of Hearing, to be published in Proc. 11th International Congress of Phon. Sci. (1987).

[3] S. G. Revoile, H. T. Bunnell, and J. M. Pickett, "On the Prediction of Phoneme Recognition by the Hearing Impaired," Symp. on Cooperative and Interactive Models of Hearing, to be published in Proc. 11th International Congress of Phon. Sci. (1987).

[4] K. N. Stevens, "Relational Properties as Perceptual Correlates of Phonetic Features," Symp. on Cooperative and Interactive Models of Hearing, to be published in Proc. 11th International Congress of Phon. Sci. (1987).

[5] W. D. Marslen-Wilson and U. H. Frauenfelder, "The Interface Between Acoustic-Phonetic and Lexical Processes," Symp. on Cooperative and Interactive Models of Hearing, to be published in Proc. 11th International Congress of Phon. Sci. (1987).

[6] A. Kohlrausch, "The Influence of Interaural Phase Uncertainty on Binaural Signal Detection," Symp. on Cooperative and Interactive Models of Hearing, to be published in Proc. 11th International Congress of Phon. Sci. (1987).