

**SPEECH COMMUNICATION**  
**6.541J-24.968J-HST710J**  
**Spring 2004**

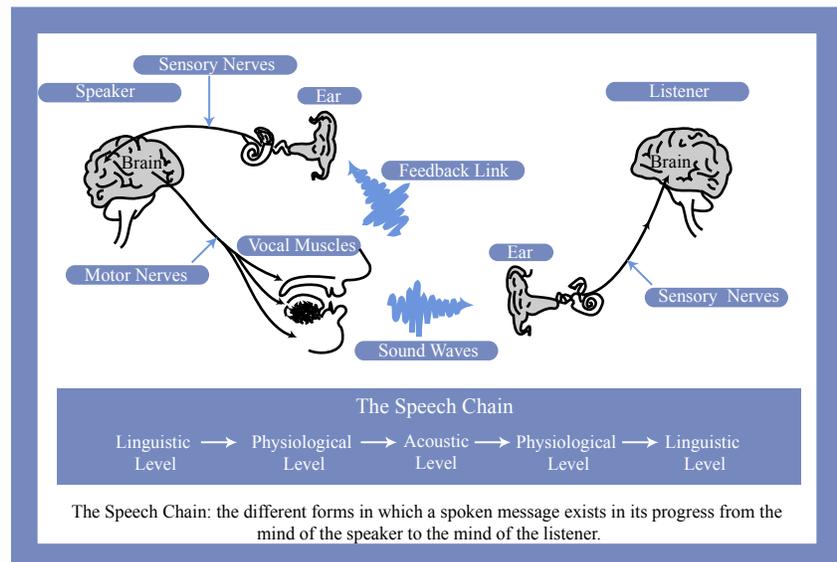
**Prof. Kenneth N. Stevens**

*Text: Stevens, K.N., **Acoustic Phonetics**, MIT Press, 1998 (paperback edition 2000).  
 (Obtainable from MIT Press in Kendall Square. This will be supplemented by additional notes.*

**Grade in this subject is calculated approximately as follows:**

Home problems	20
Lab reports	20
Quizzes	30
Report on project	<u>30</u>
	<u>100</u>

**Overview of subject**



**The Speech Chain**

**Linguistic level**

Underlying the production of speech are words, which are organized into sentences. Words, in turn, are composed of syllables, which are constructed from phonemes (or segments), so that a word like "cat" has three such segments. The segments, in turn, can be classified in terms of discrete, binary features. Thus **p, t, k**, have in common that they have the feature **voiceless**, whereas **b, d, g** are **voiced**, etc. All six of these consonants

have a common attribute that they are **stop** consonants. At this level, then, the representation of language is digital. The word sequences are organized according to syntactic constraints.

#### Physiological level

The segments and features that make up words specify instructions indicating how the sounds in the words are produced by the articulators --- the tongue, lips, larynx, etc. These instructions state how the different articulators are to be moved and coordinated via the muscles, and, for a given feature, there may be some instructions that depend on the context in which it appears. There are limitations in the speed at which the articulators can be moved. A central problem is understanding the transformation from the discrete linguistic representation to the continuous movements of the articulators.

#### Acoustic level

The movements of the articulators, together with the movement of air from the lungs, results in the creation of sound sources which are filtered by acoustic cavities that are shaped by the articulators. The physics of sound production draws on concepts of acoustics, source-filter theory, and signal processing. The sound radiated from a speaker's lips is the most accessible component of the speech chain. The sound has some quantal aspects and some continuous aspects. This sound impinges on the ear of a listener where it is converted to mechanical vibration of the eardrum.

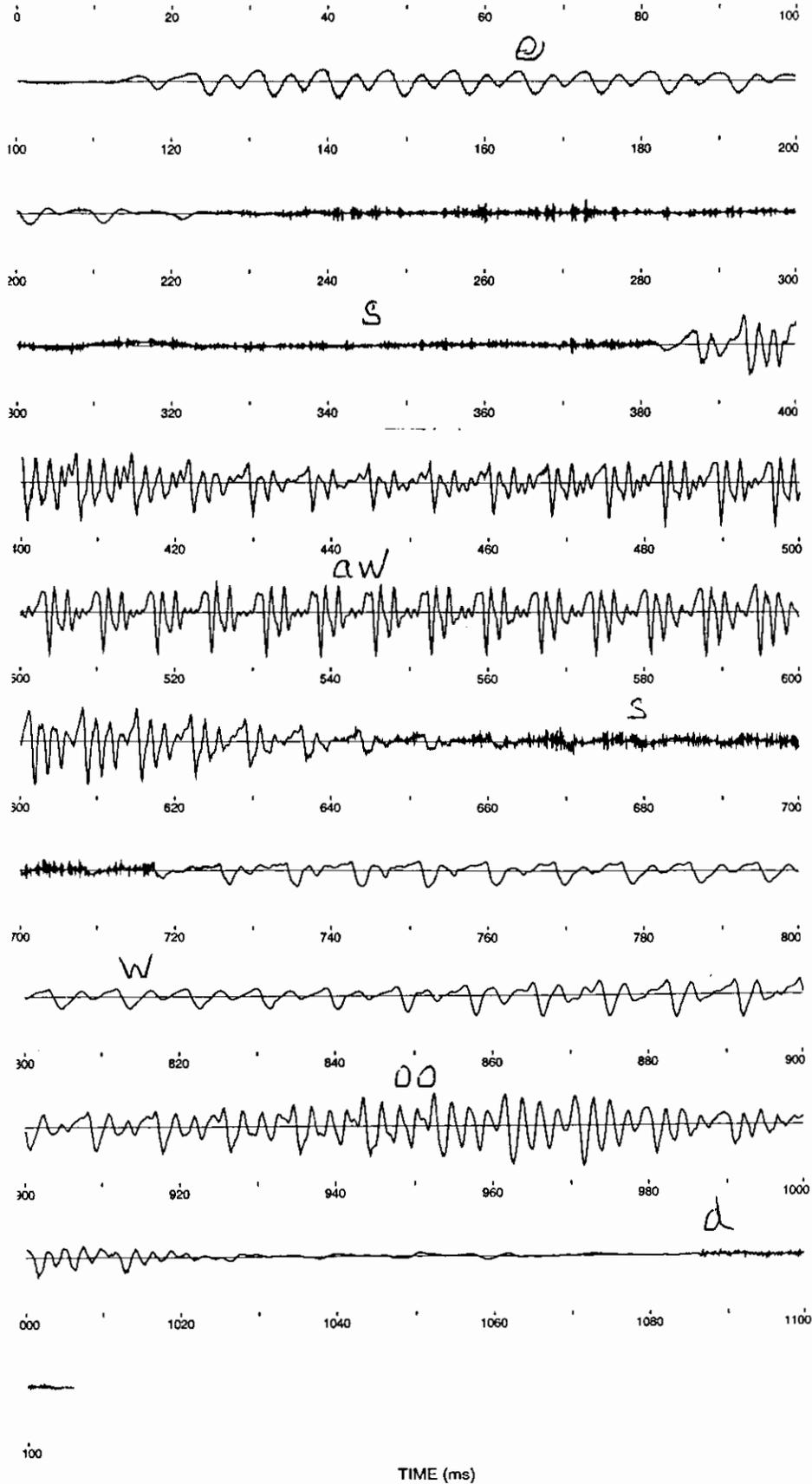
#### Auditory physiology

In a series of stages, the sound is converted to mechanical motion, it undergoes some frequency analysis in the ear, and is converted into electrical activity in the auditory nerve and in higher centers in the brain. The listener must decode this sound in order to extract the intended sequence of words at the linguistic level. The decoding presumably requires identification or estimation of the segments and features that underlie the sound pattern.

# The speech waveform

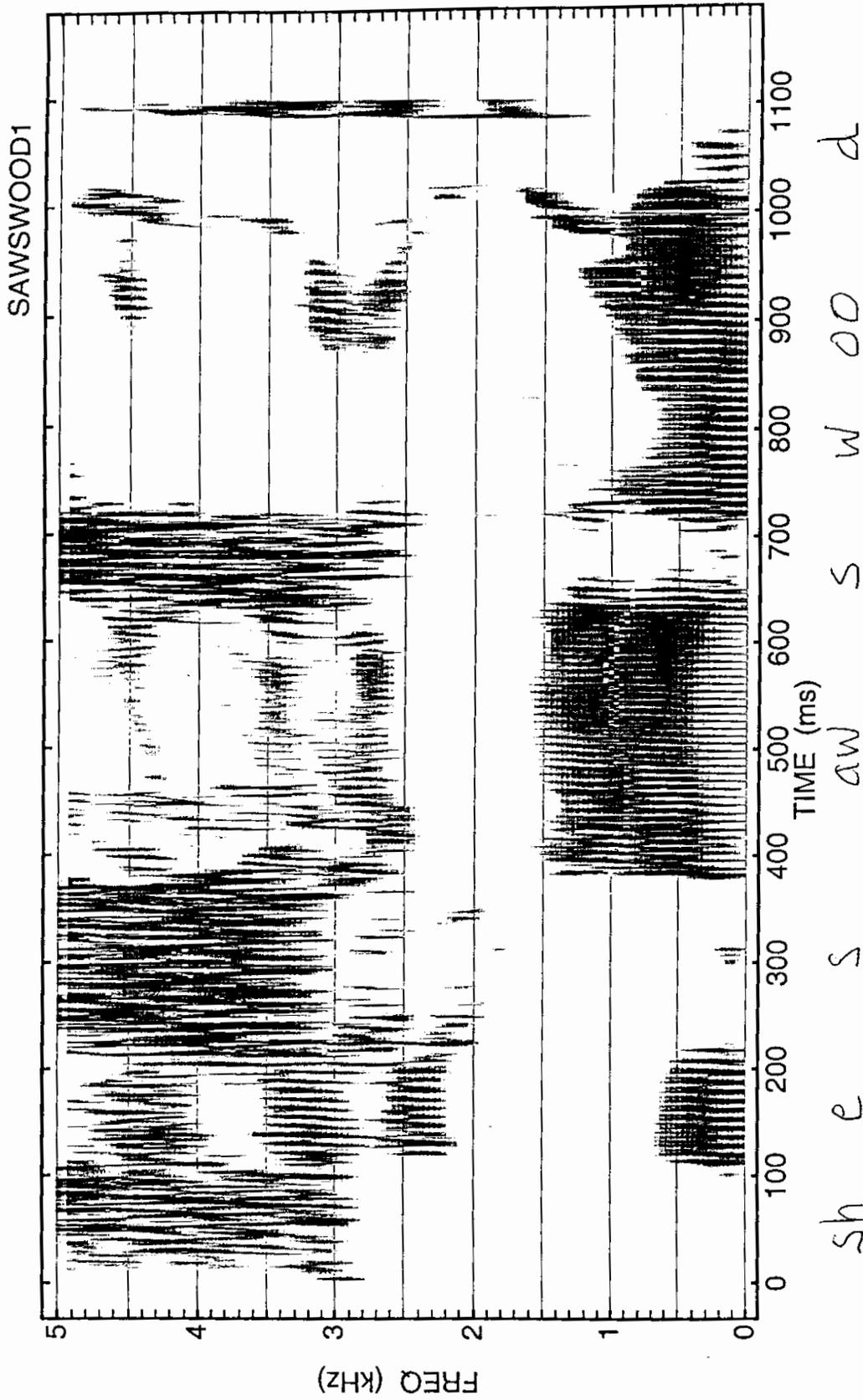
sh

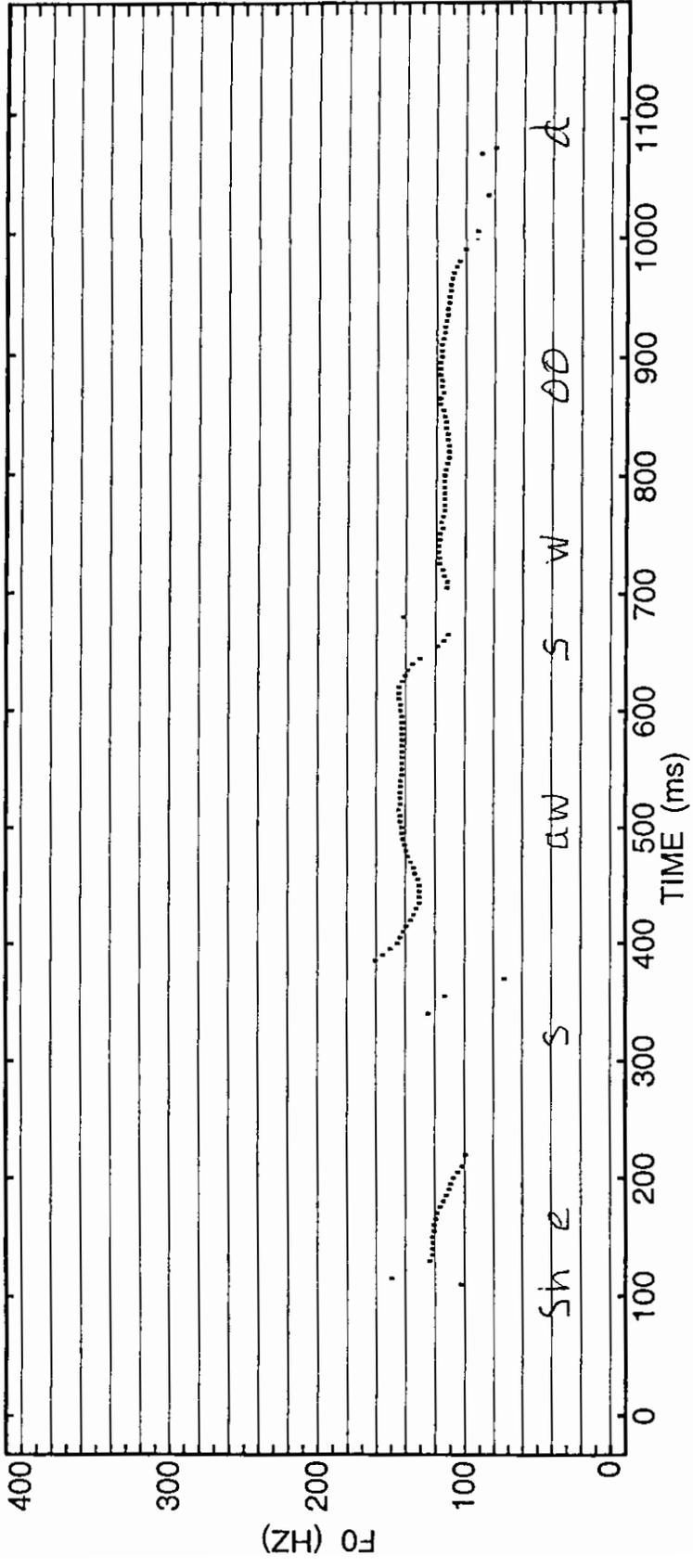
3



She saws wood

A spectrogram





sh e wa d