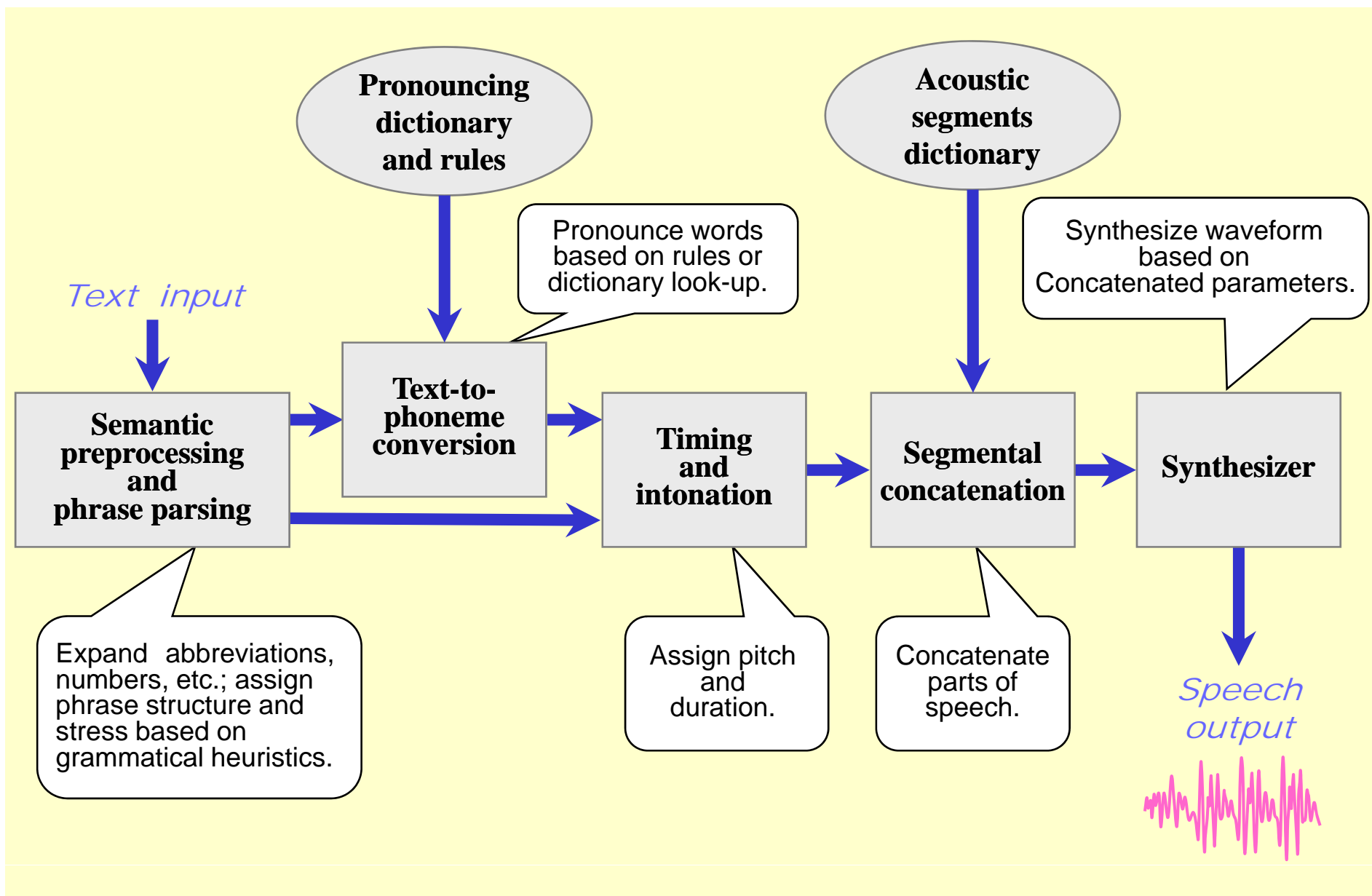


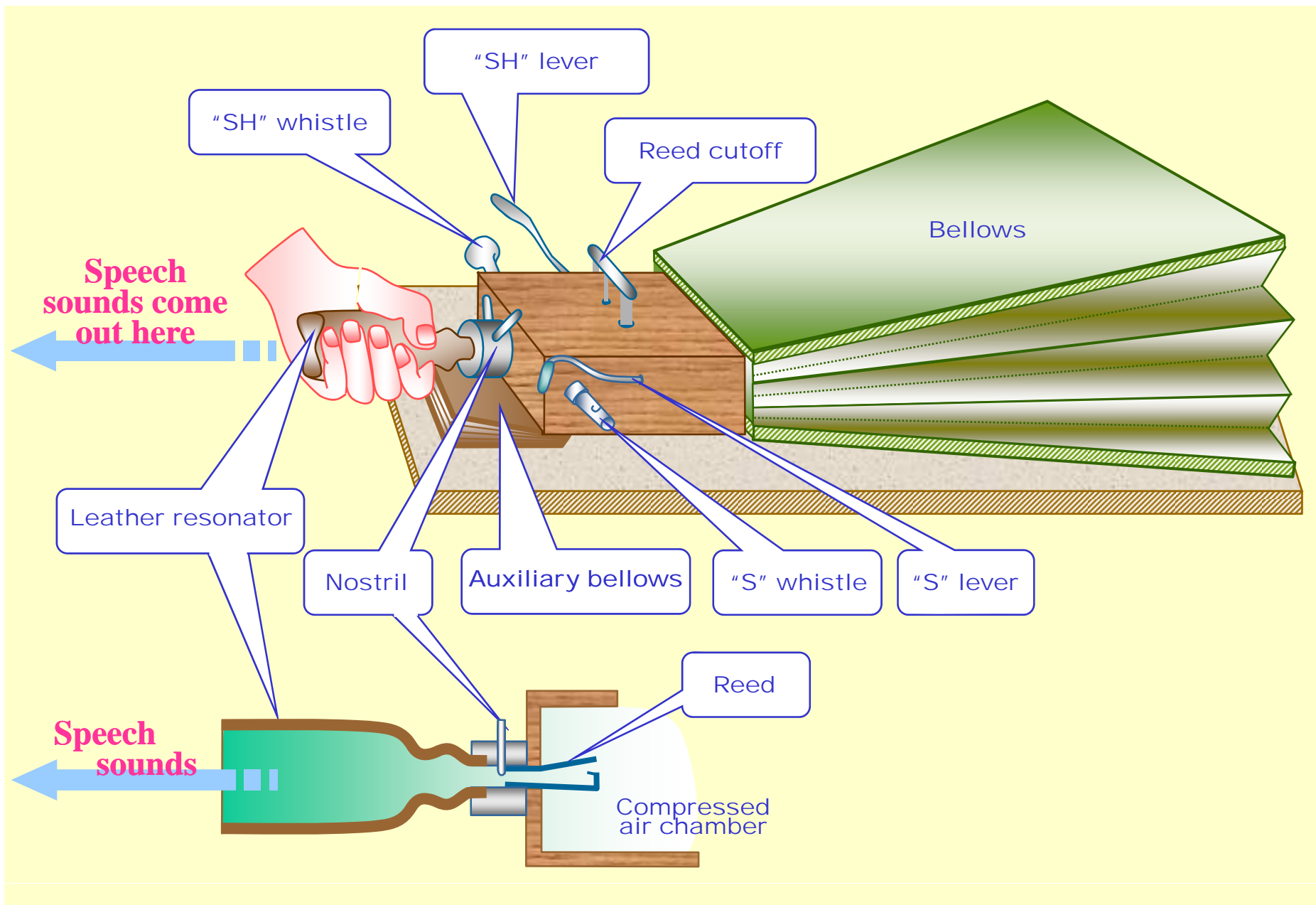
Speech Synthesis

Sadaoki Furui

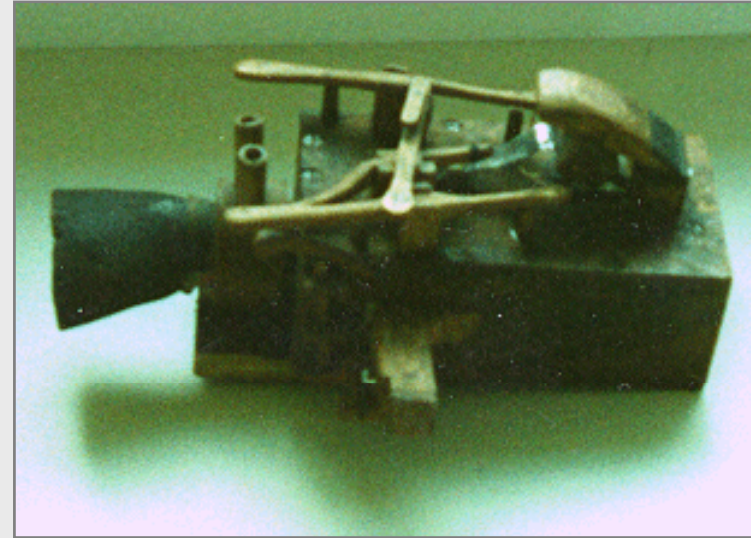
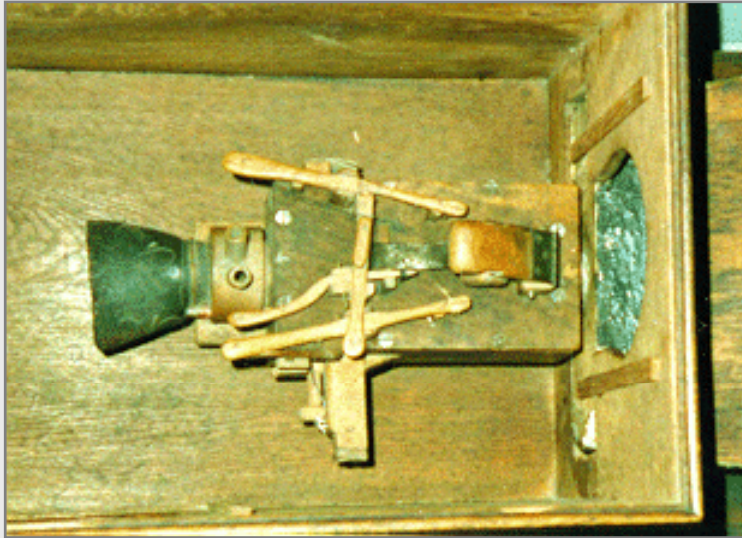
Tokyo Institute of Technology
Department of Computer Science
furui@cs.titech.ac.jp



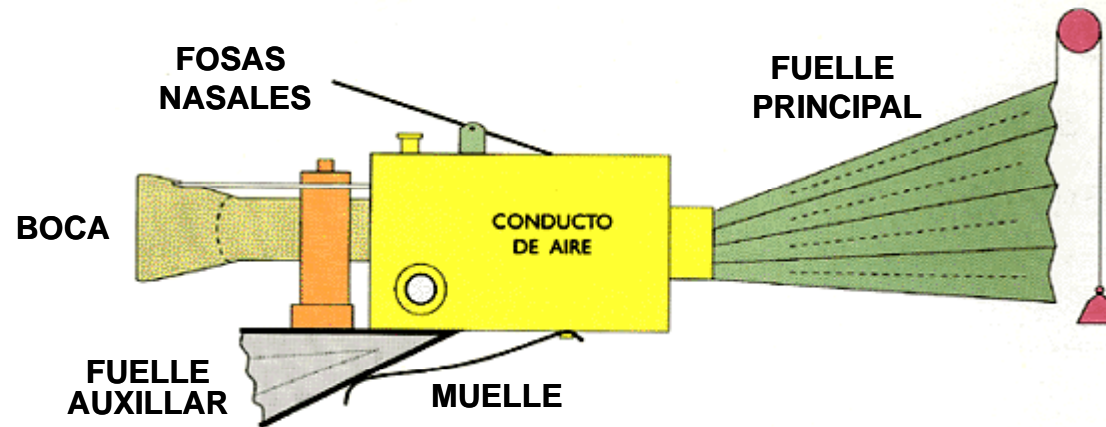
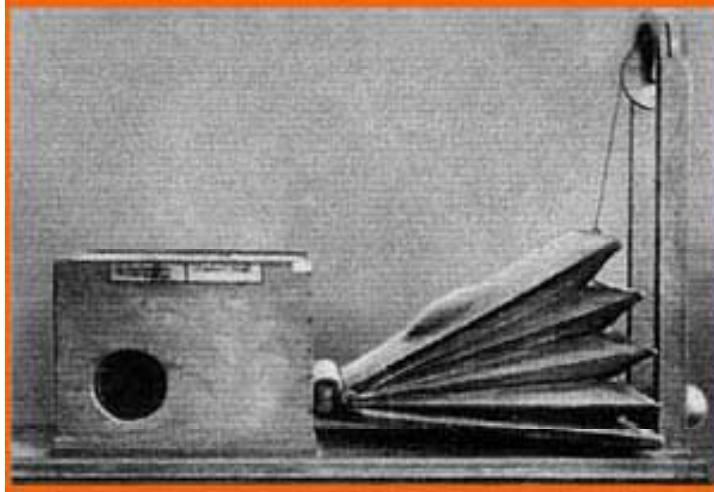
Principal elements of text-to-speech conversion system



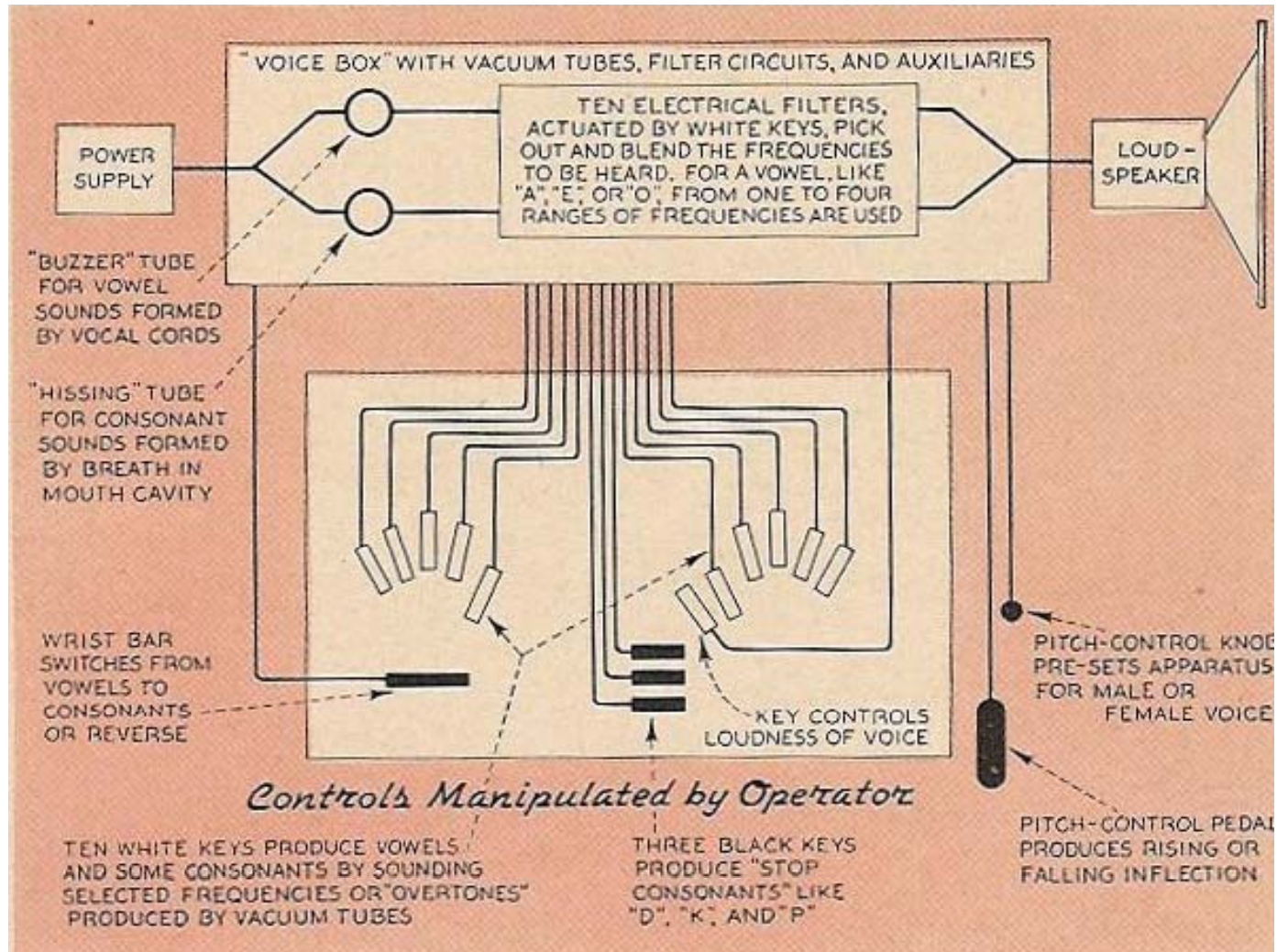
Mechanical speech synthesizer by von Kempelen



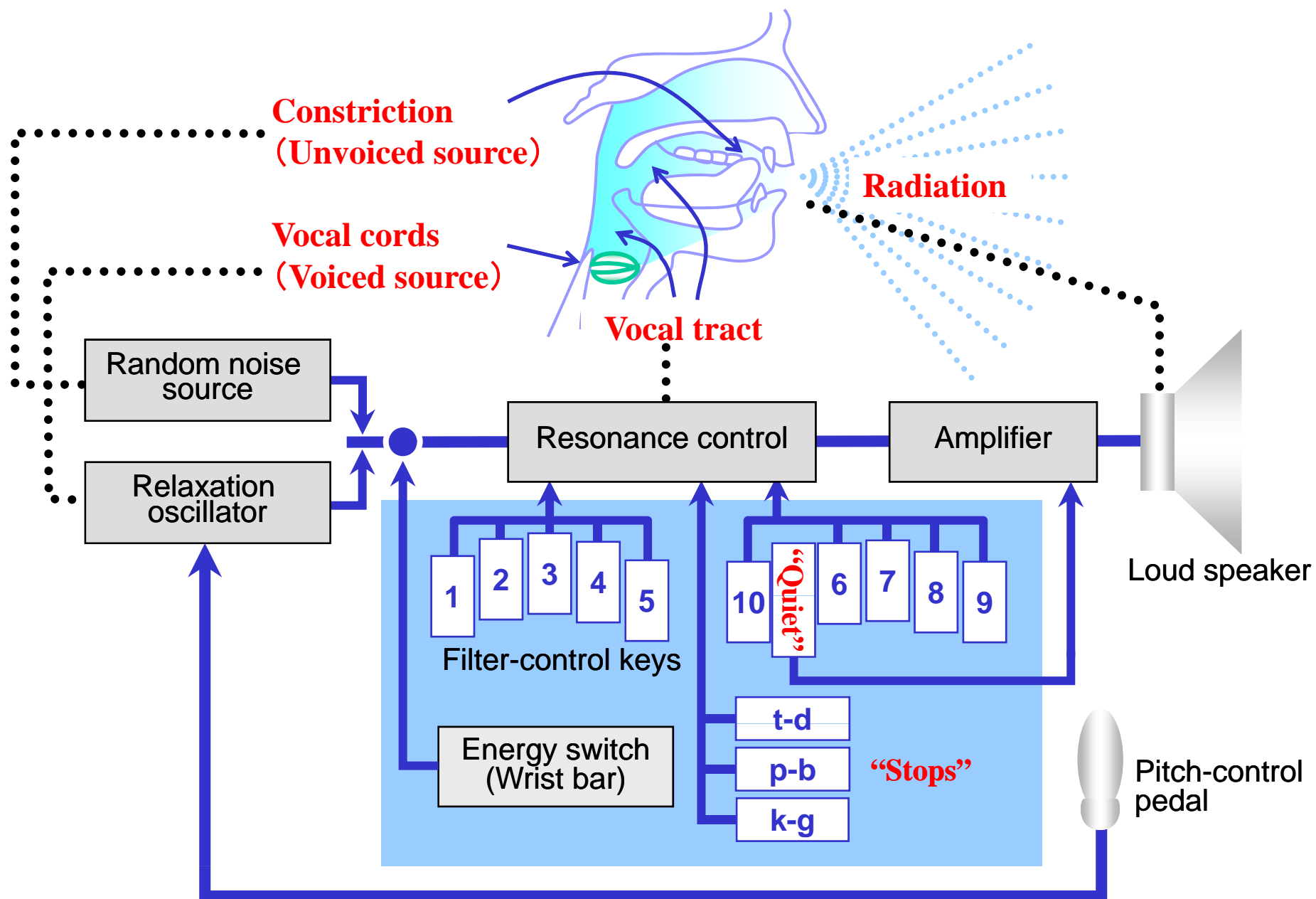
**The sound production mechanism of
Kempelen's speaking machine.**



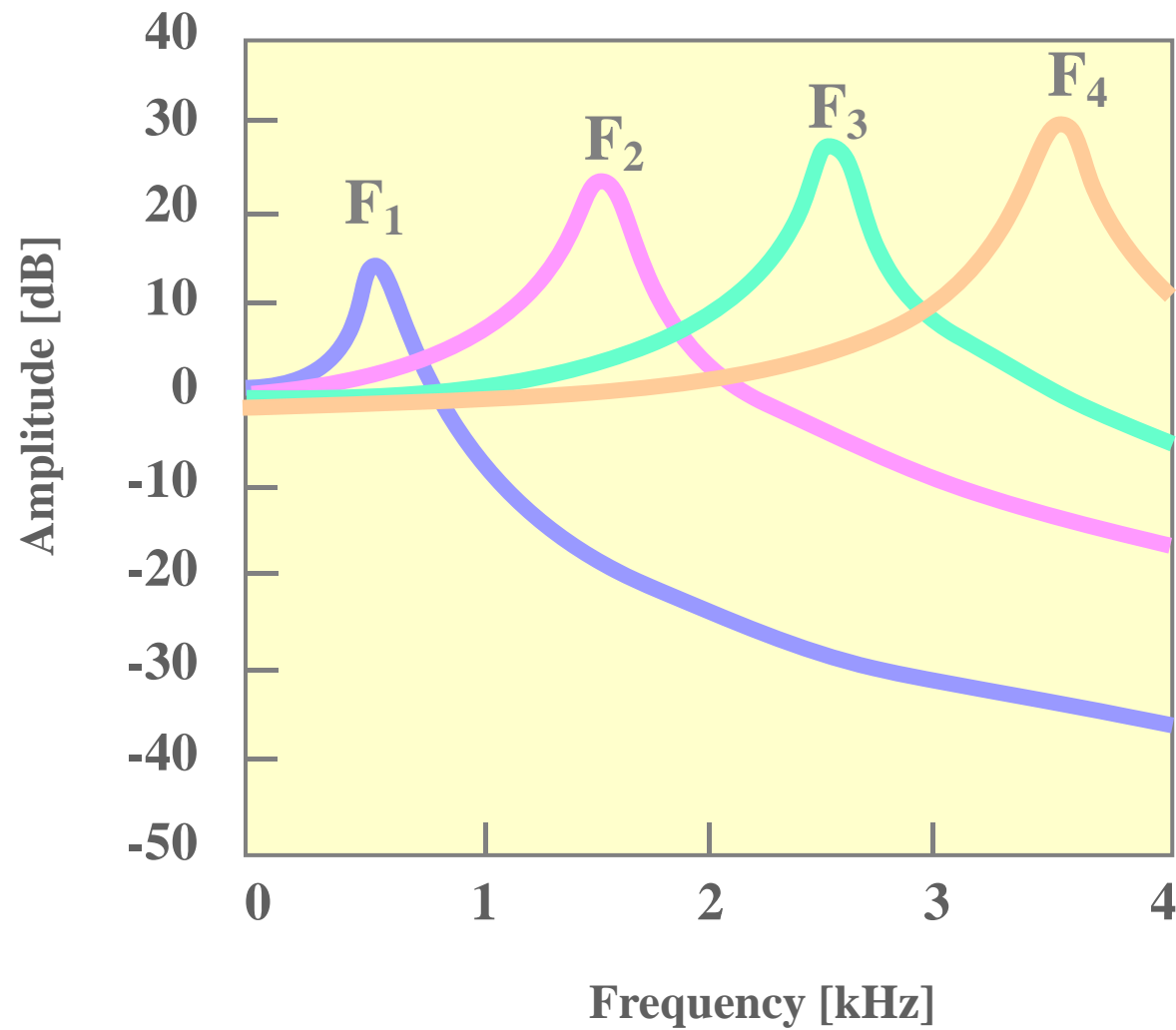
Von Kempelen's speaking machine, as it can be seen in the Deutsches Museum in Munich, and seen from above, with the cover of the box



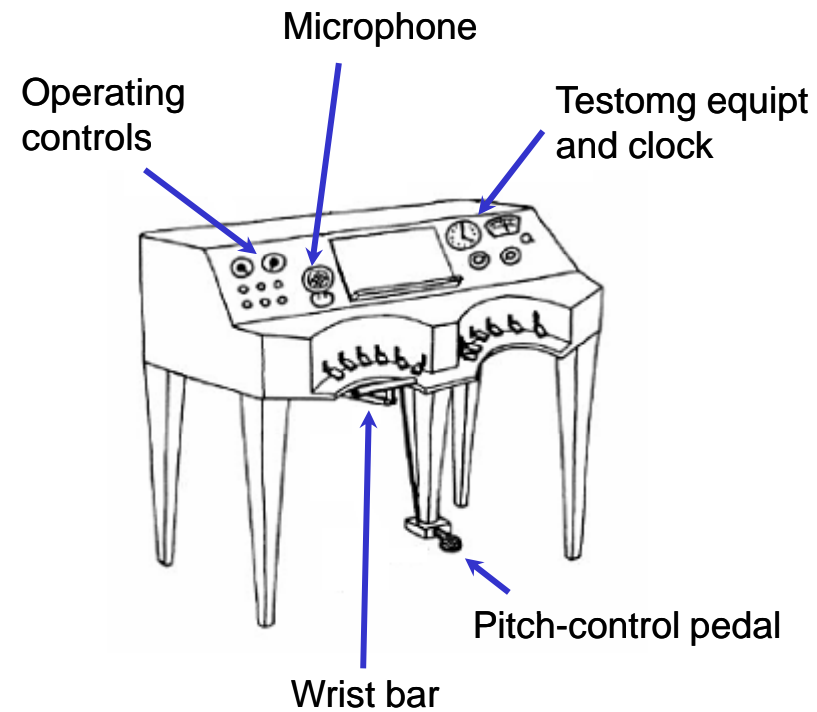
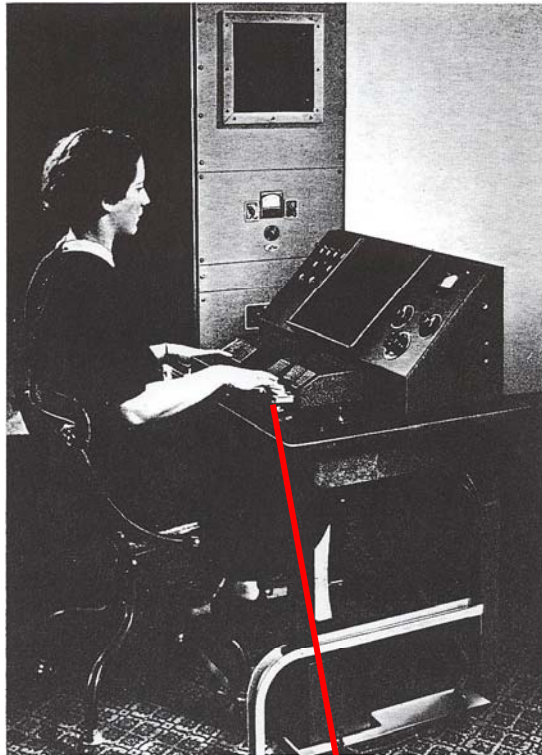
Voder synthesizer (1939)



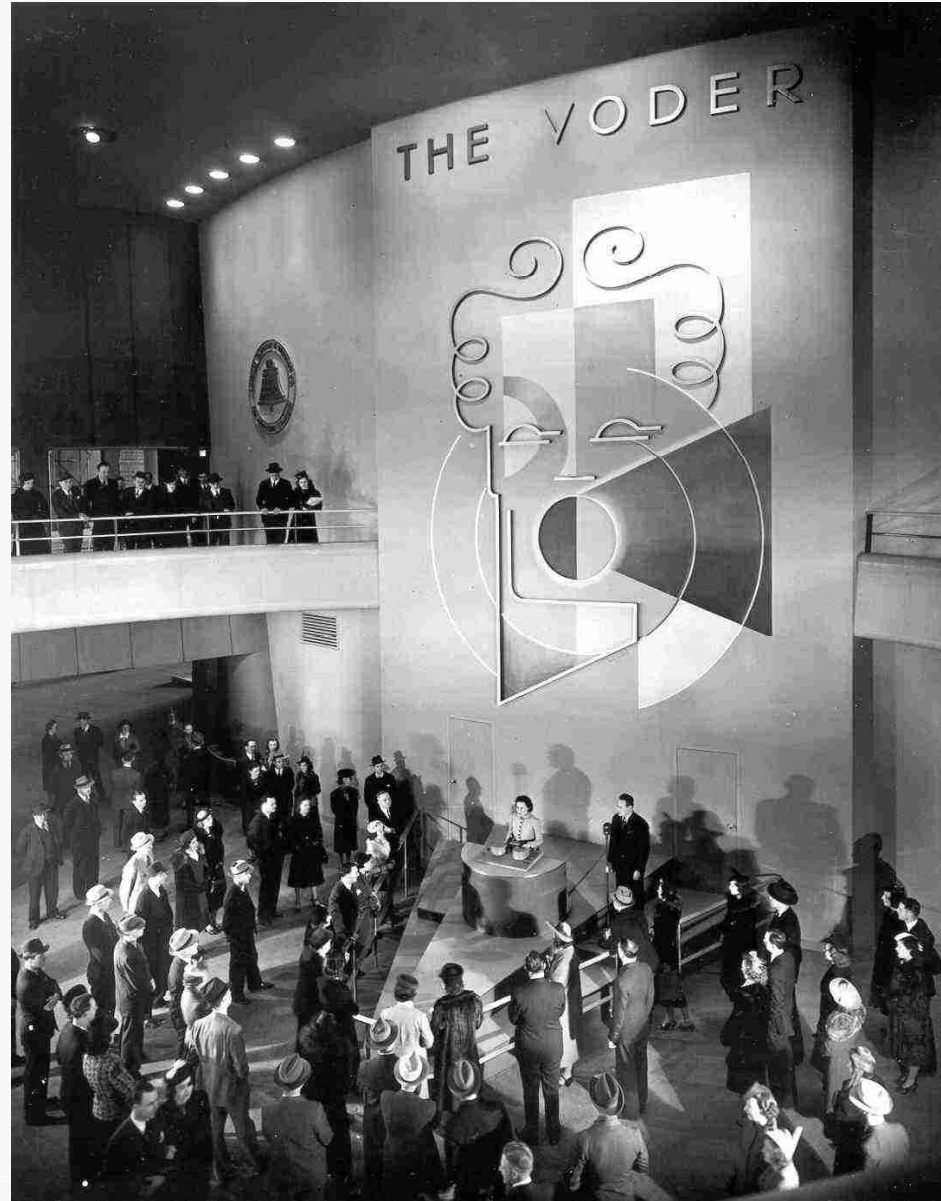
Voder synthesizer



Contribution of each formant to the amplitude spectrum














The voder as demonstrated by Mrs. Harper at the Franklin institute



The voder being demonstrated at the New York world's fair

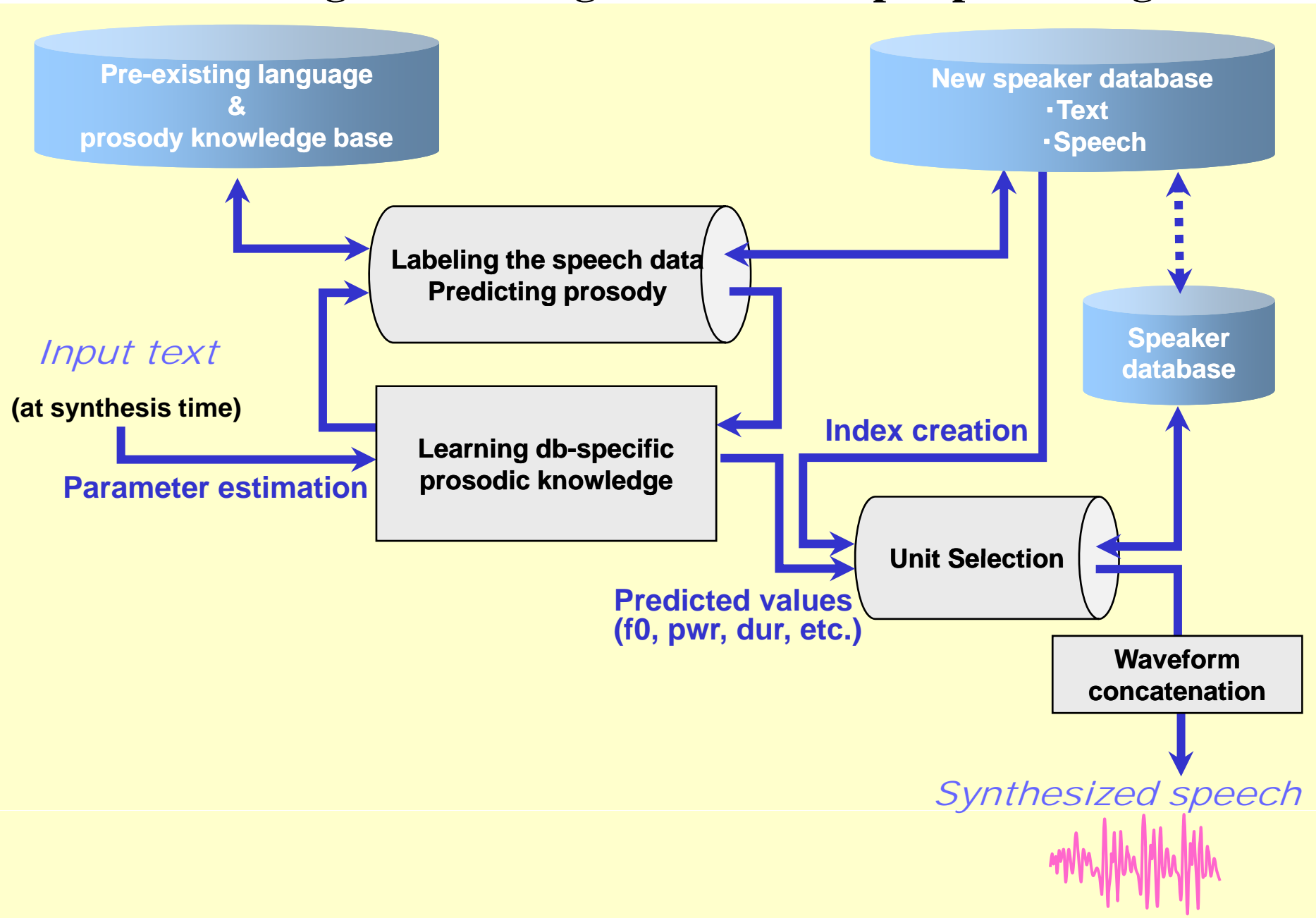
History of speech synthesis

1	The VODER of Homer Dudley 	1939
11	The DAVO articulatory synthesizer developed by George Rosen at M.I.T. 	1958
6	Copying a natural sentence using the second generation of Gunnar Fant's OVE cascade formant synthesizer 	1962
13	Linear-prediction analysis and resynthesis of speech at a low-bit rate in the Texas Instruments Speak-'n-Spell toy, Richard Wiggins	1980
30	The M.I.T. MITalk system by Jonathan Allen, Sheri Hunnicutt, and Dennis Klatt 	1979
33	The Klattalk system by Dennis Klatt of M.I.T. which formed the basis for Digital Equipment Corporation's DEC-talk commercial system 	1983
35	Several of the DECtalk voices     	
36	DECtalk speaking at about 300 words/minute 	

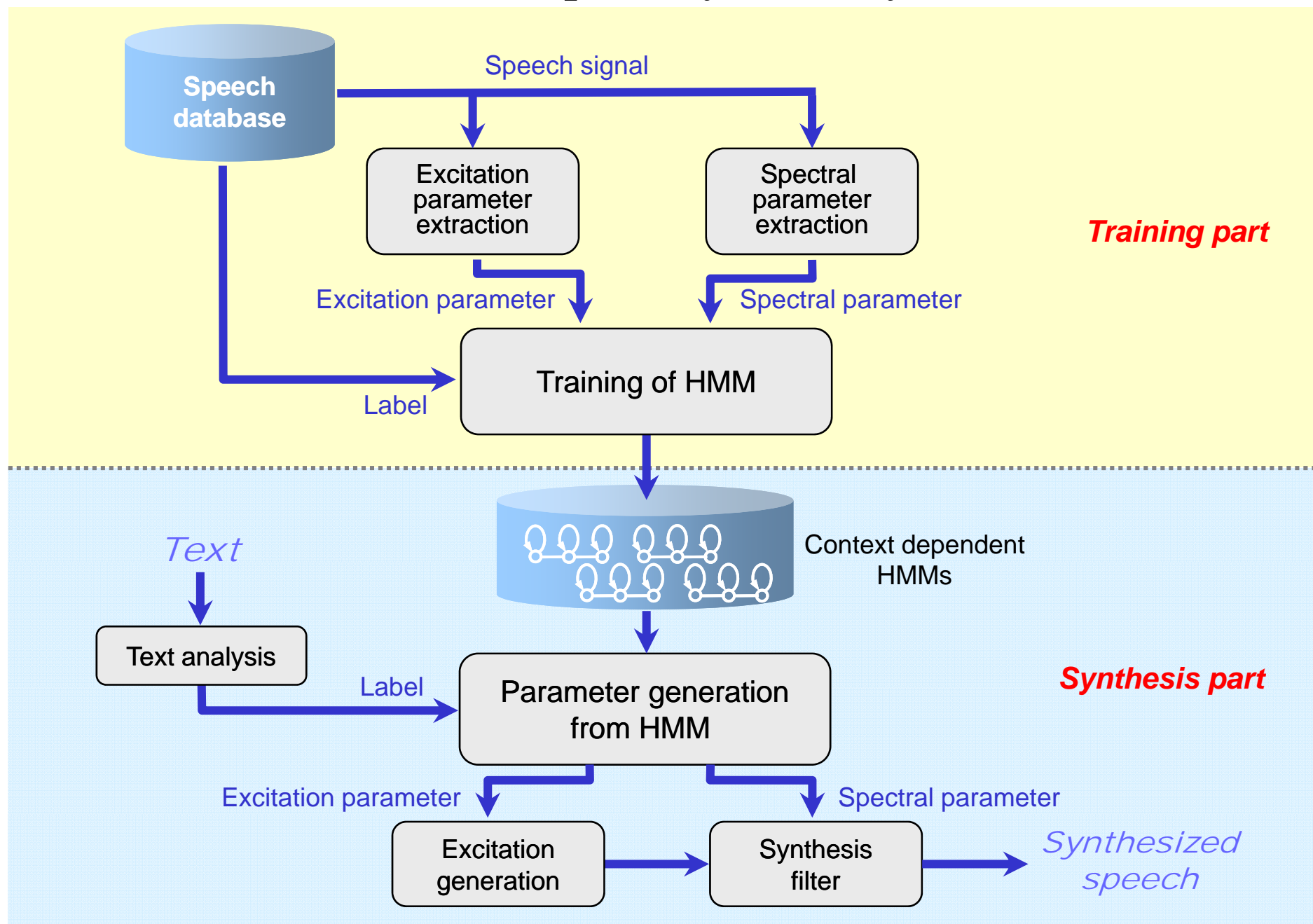


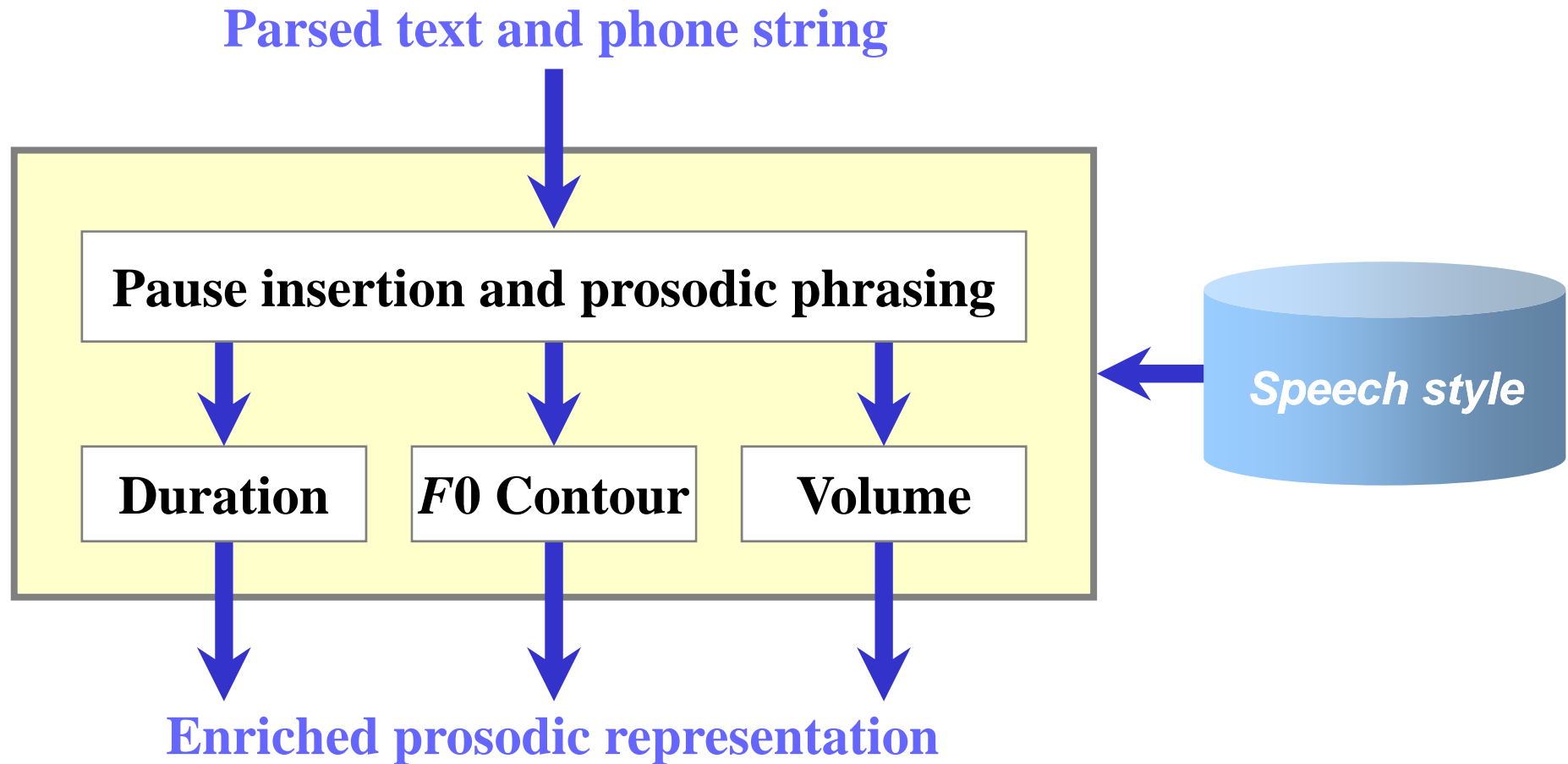
Speak-'n-Spell toy

Flow diagram showing CHATR's corpus processing

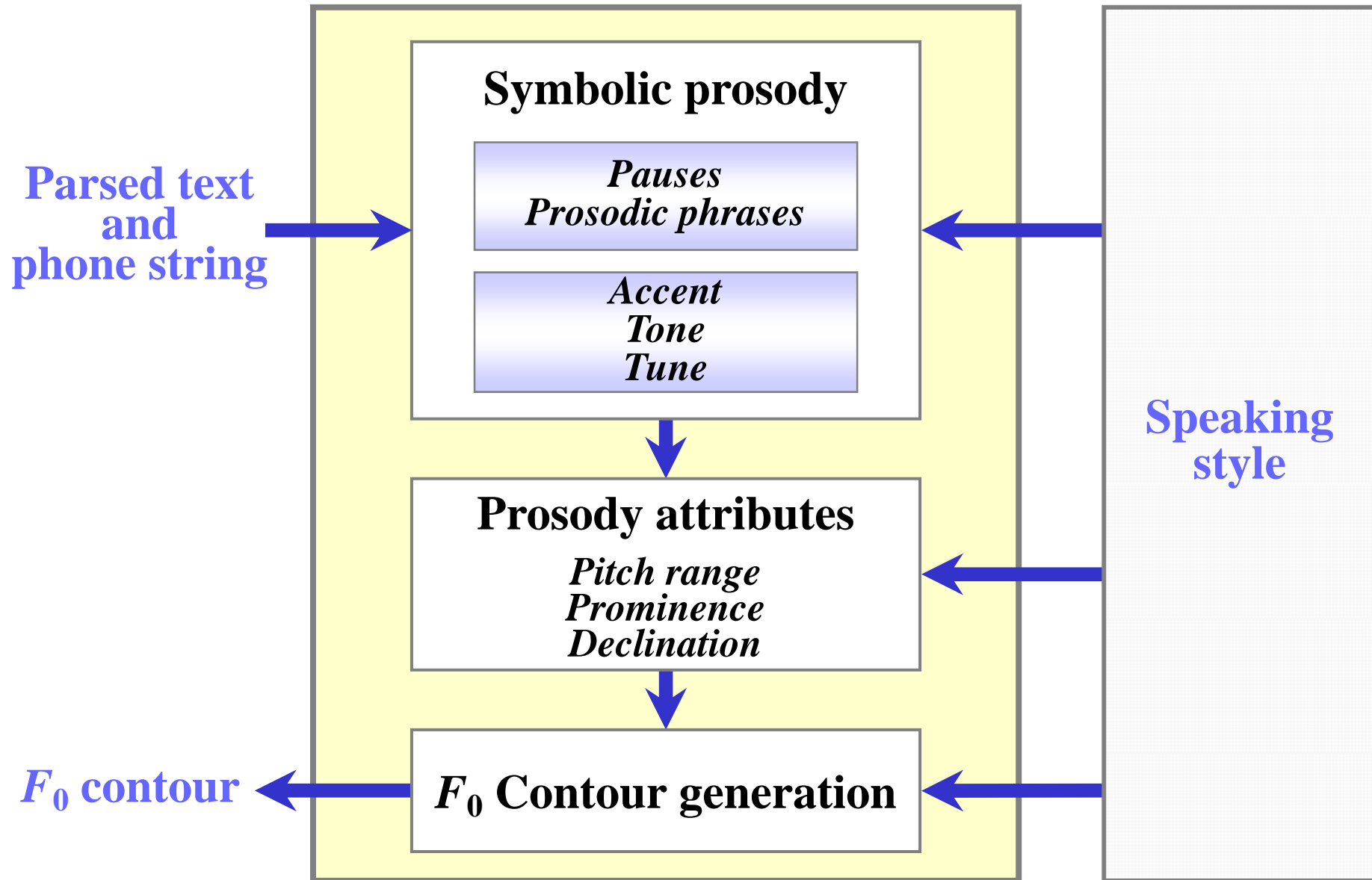


HMM-based speech synthesis system



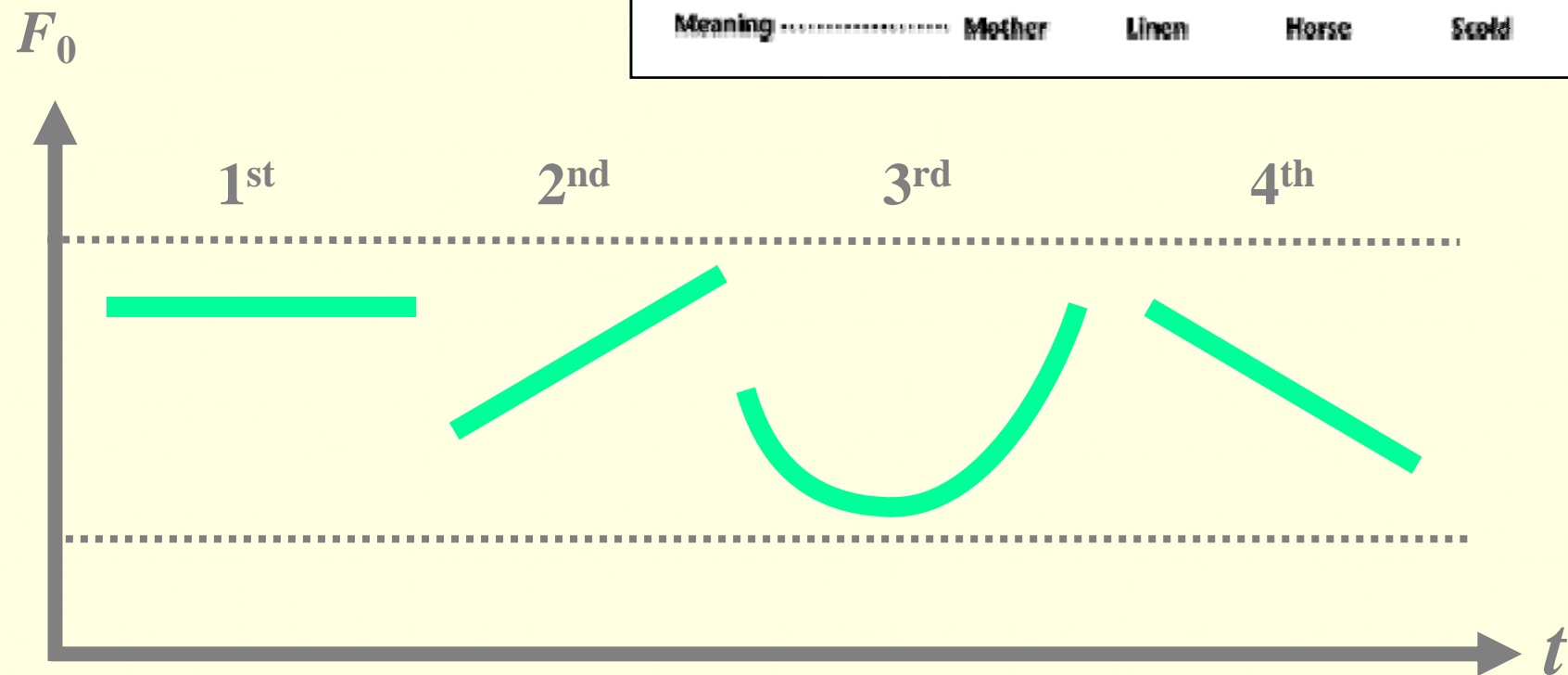


Block diagram of a prosody generation system; different prosodic representations are obtained depending on the speaking style we use.



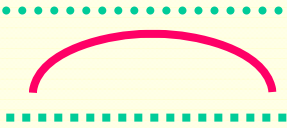




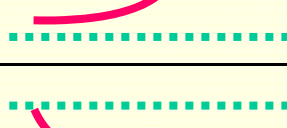
Pitch generation decomposed in symbolic and phonetic prosody

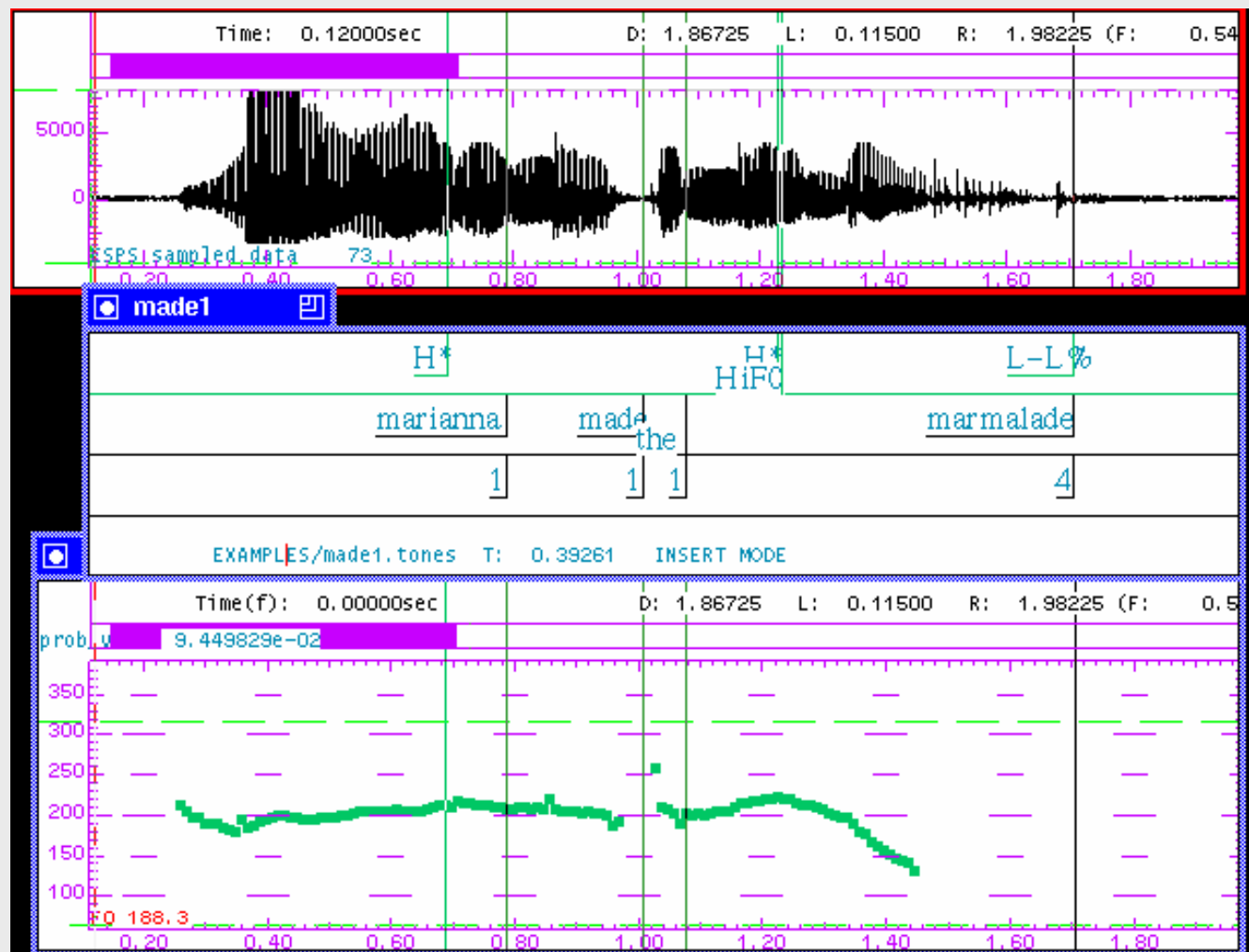
Tone number	1st tone	2nd tone	3rd tone	4th tone
Tone symbol	[-]	[/]	[ˇ]	[\]
Voice pitch	High Low	High Low	High Low	High Low
Syllable	mā (妈)	má (麻)	mǎ (马)	mà (骂)
Meaning	Mother	Linen	Horse	Scold



The four Chinese tones

ToBI pitch accent tones

ToBI tone	Description	Graph
H*	<i>Peak accent</i> — a tone target on an accented syllable which is in the upper part of the speaker's pitch range.	
L*	<i>Low accent</i> — a tone target on an accented syllable which is in the lowest part of the speaker's pitch range.	
L*+H	<i>Scooped accent</i> — a low tone target on an accented syllable which is immediately followed by a relatively sharp rise to a peak in the upper part of the speaker's pitch range.	
L*+!H	<i>Scooped downstep accent</i> — a low tone target on an accented syllable which is immediately followed by a relatively flat rise to a downstep peak.	
L+H*	<i>Rising peak accent</i> — a high peak target on an accented syllable which is immediately preceded by a relatively sharp rise from a valley in the lowest part of the speaker's pitch range.	
!H*	<i>Downstep high tone</i> — a clear step down onto an accented syllable from a high pitch which itself cannot be accounted for by an H phrasal tone ending the preceding phrase or by a preceding H pitch accent in the same phrase.	



“*Marianna made the marmalade*”, with an H* accent on *Marianna* and *marmalade*, and final L-L% marking the characteristic sentence-final pitch drop. Note the use of 1 for the weak inter-word breaks, and 4 for the sentence-final break (after Beckman)