

Combining different acoustic streams - scaling from small to large tasks

Morgan

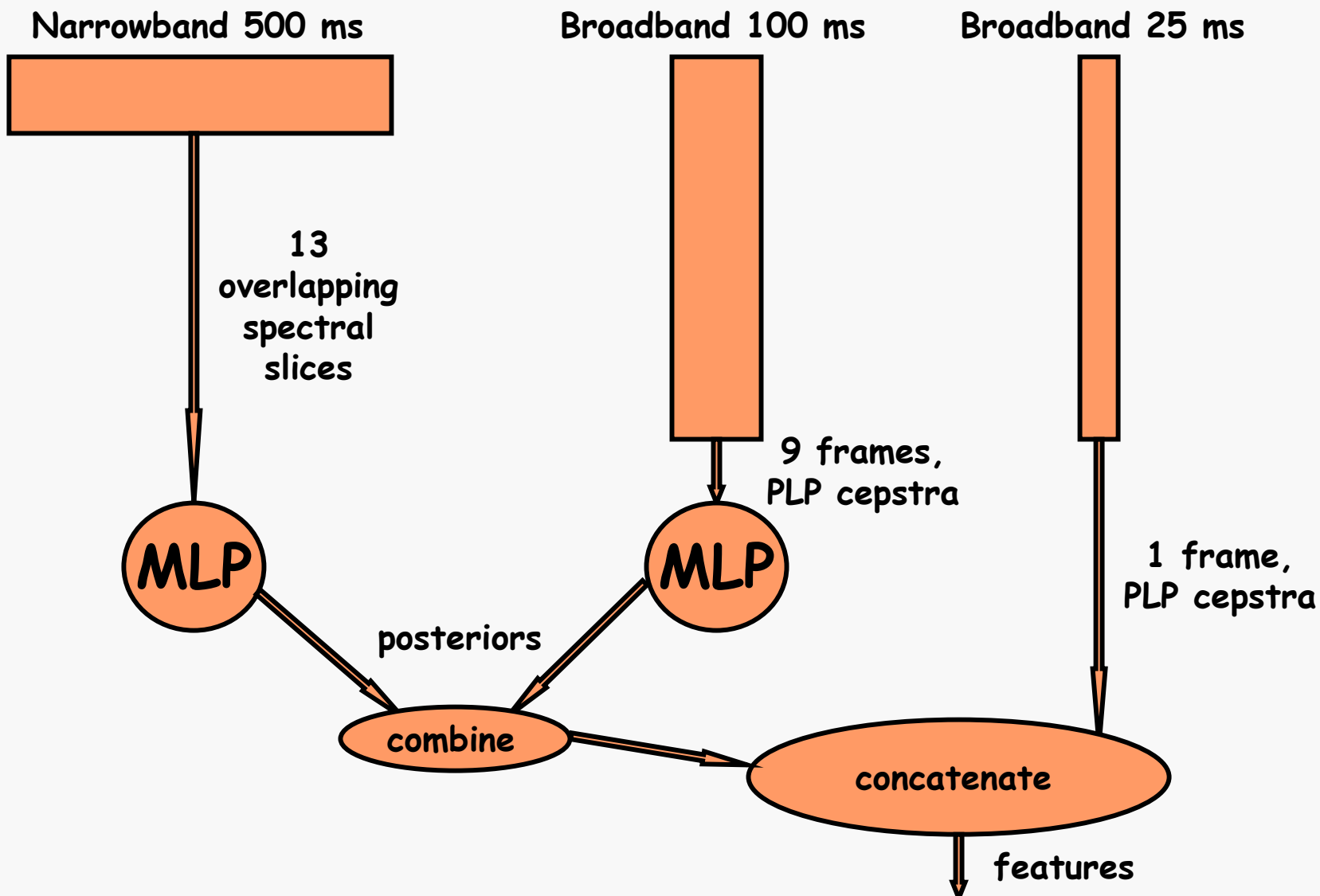
ICSI/UC Berkeley

(representing work from
ICSI/SRI/IDIAP/Columbia/UW)

"Novel Approaches" to ASR (EARS)

- Multiple front ends/models work better
 - analogy to multimodal
 - analogy to multiple mappings in cortex
 - specific case of incorporating TRAP/Tandem
(primarily temporal vs spectral)
- Test case for how to scale up from small tasks
 - providing new features for a "big" system (SRI)
 - New results on conversational speech recognition

Multiple scales in Time-Frequency



Experimental WER

(+ % relative reduction) Results

	Jan	May	June	Late Aug
Front end	Numbers	CTS 500 (w/23 hrs training)	Eval 2001 - (w/ 64 hrs training)	Eval 2001 - (w/vtln, phone loop adaptation, better norm, LM)
PLP baseline	4.0	43.8	43.8	37.1
PLP-TA-TR1	3.2 (20)	39.5 (9.9)		
PLP-TA-TR3		39.1 (10.8)	40.5 (7.6)	33.9 (8.7)
PLP-TA-HAT1				33.2 (10.6)

TA-TR1 = Tandem and 1-band TRAPs, log posterior combo

TA-TR3 = Same, but with 3-band TRAPs

TA-HAT1 = 1-band but using Hidden Activation TRAPS

My last slide

- These techniques provide even more margin in noise
- More well-chosen streams should be better
- Still more to learn about the combination/modeling
- Taught us something about scaling from small tasks (quick turnaround) to large ones
 - use multiple steps
 - have people working on all levels