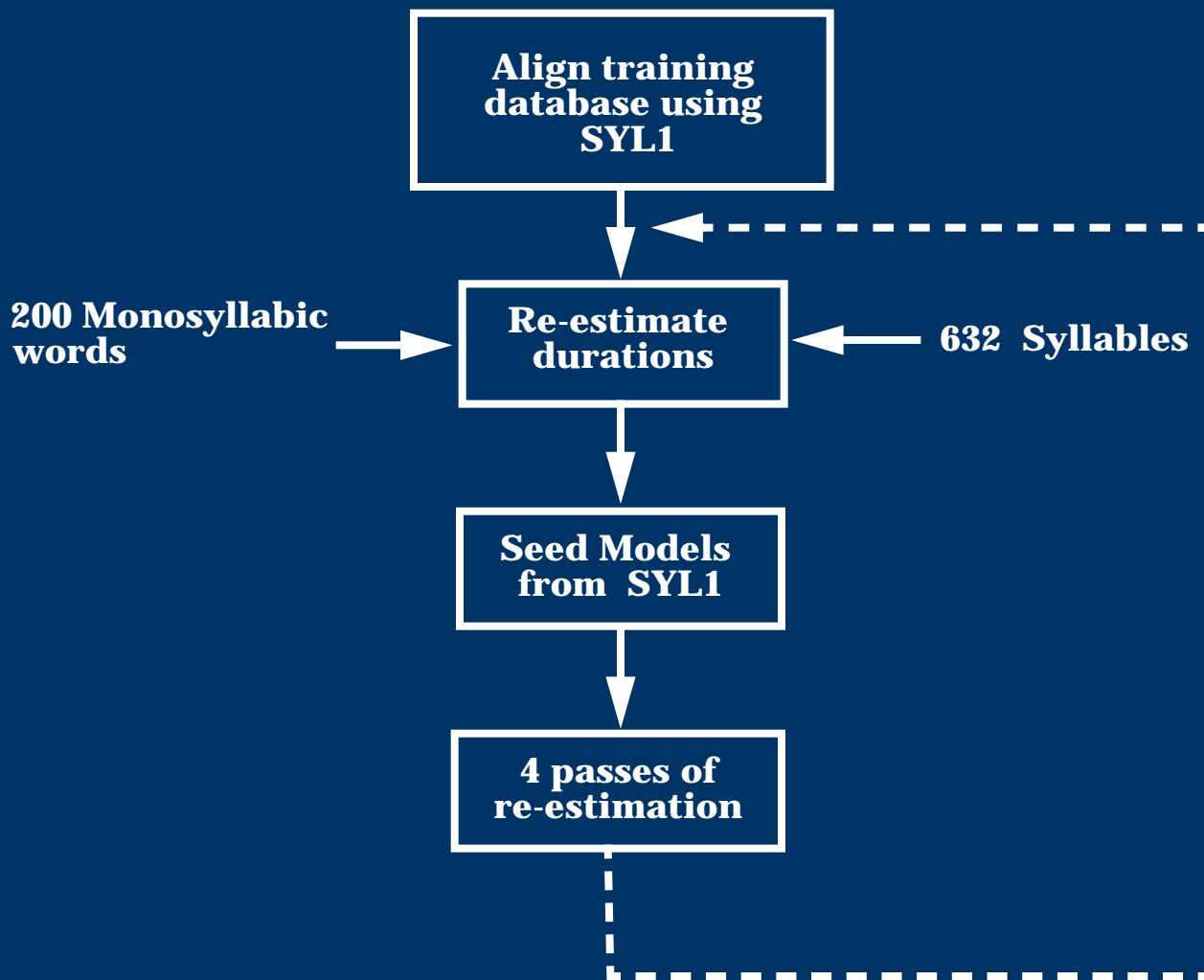
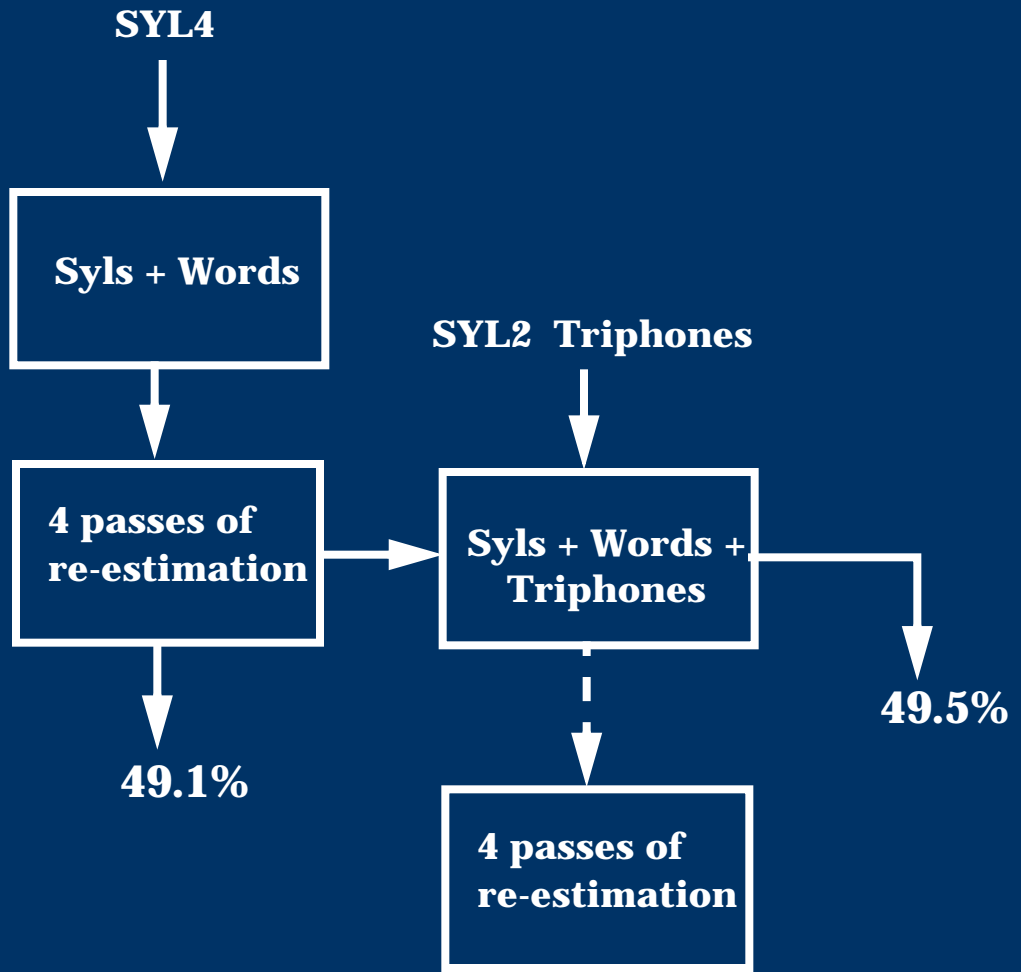


$$\mathbf{N} = \mathbf{E}[\mathbf{S1}] + 2 \text{stddev}(\mathbf{S1}) = \mathbf{f}(\mathbf{p})$$

S1: duration of state s1 in frames





_ae_t aa _m_ih_k → _ae_t t-aa+m _m_ih_k

9023 Syllables Flatstarted



**Trained to
8 gaussians/state**



**800 most frequent
syllables**



Monophones



Test system



55.1%

Monophones



2419 Frequent syllables



**Flat-start models
Models trained to 8 gaussians/state**



WER 54.6% (S 36.9, D 13.2, I 4.6)

**Baseline word-internal
triphones**

**2419 Syllables from
hybrid system**



4 passes of re-estimation



WER 52.9% (S 37.2, D 11.5, I 4.3)

**Baseline word-internal
triphones**

**800 Syllables from
hybrid system**

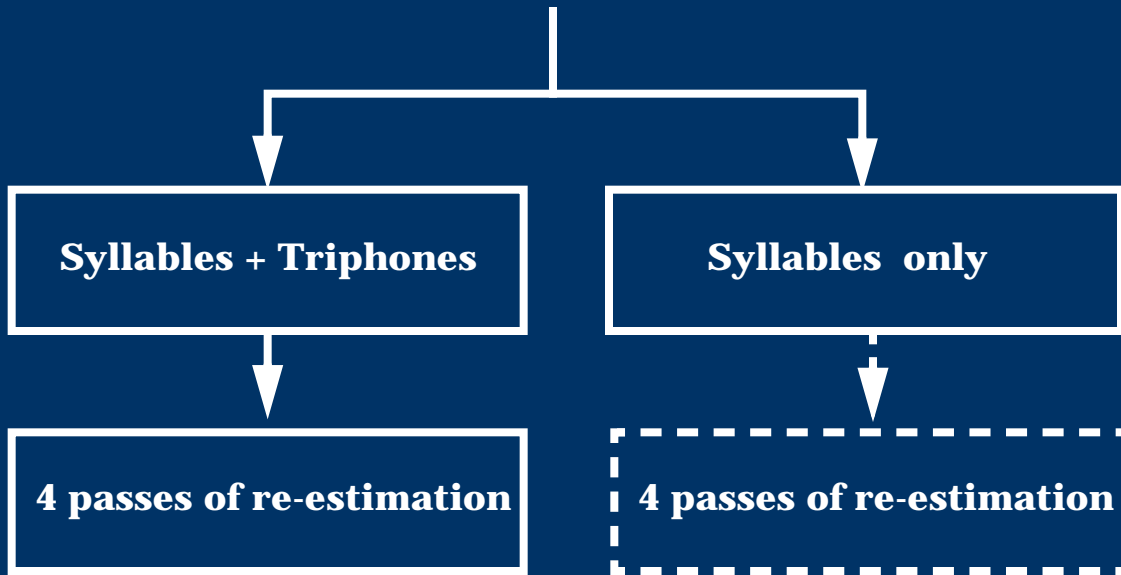


4 passes of re-estimation

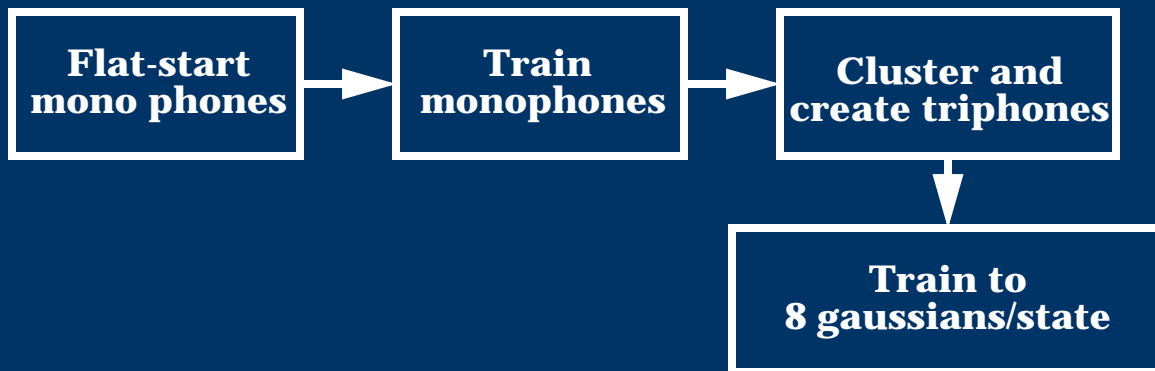


WER 51.7% (S 33.9, D 14.3, I 3.5)

**800 Syllables + Triphones
from hybrid system**

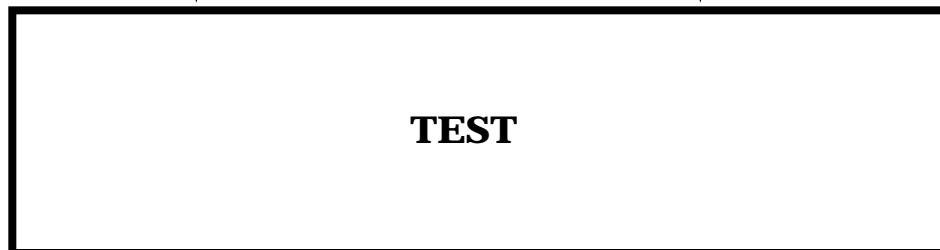


WER 49.9% (S 32.3, D 14.1, I 3.5)



Monophones

800 Frequent syllables



TEST



WER 55.1% (S 35.7, D 16.9, I 2.5)