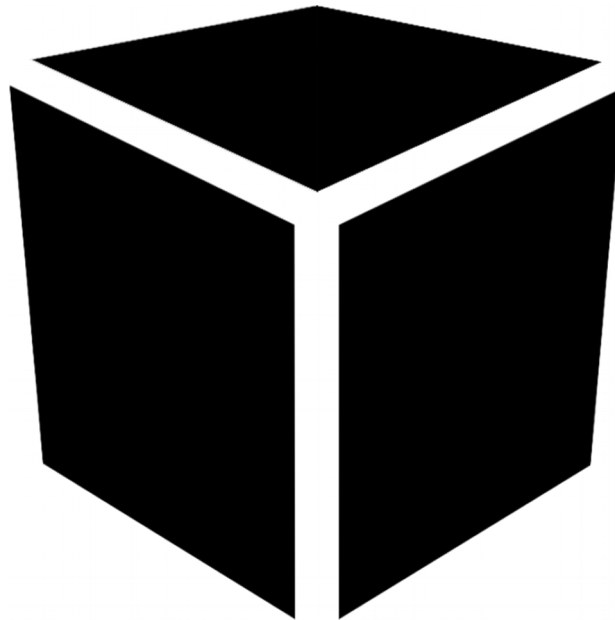
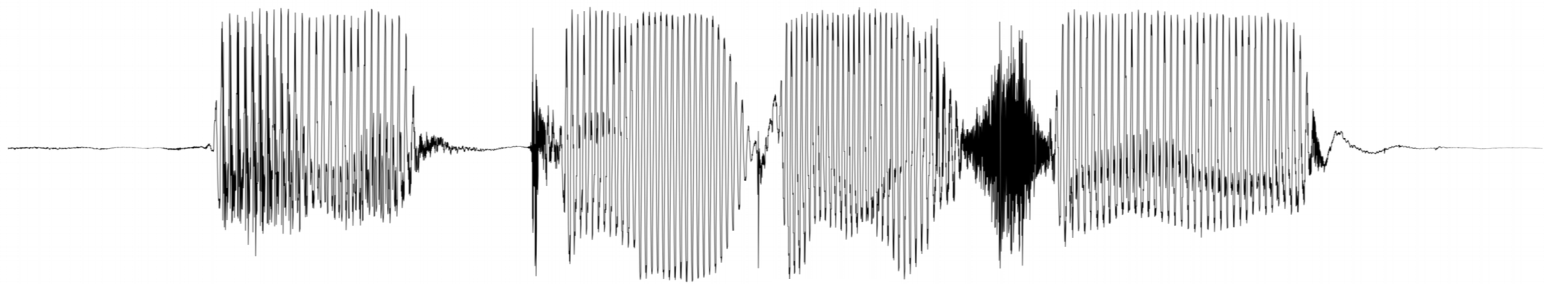


Automatic Speech Recognition:

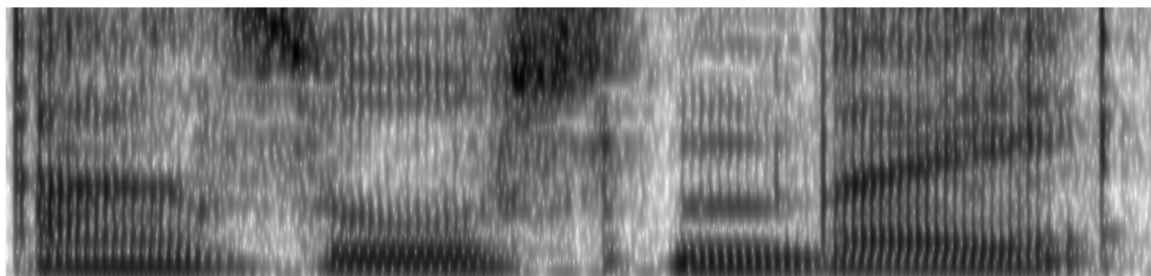
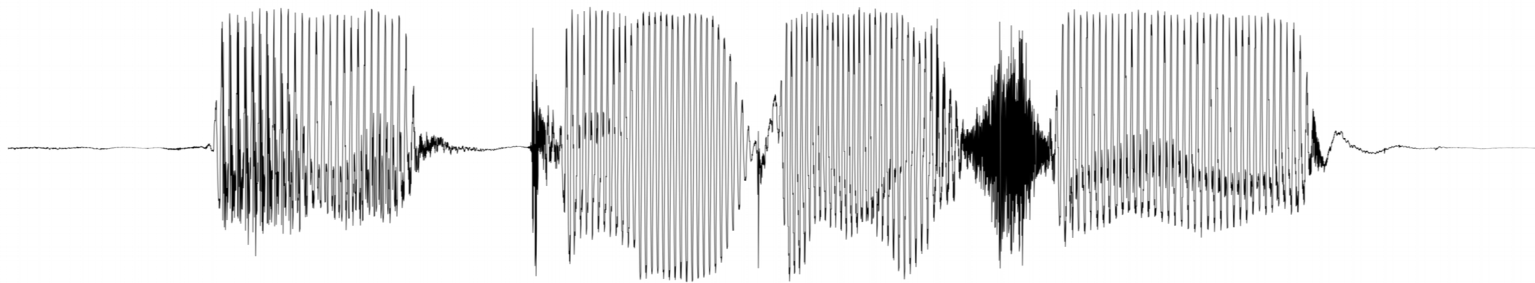
Gaussians and
Neural Nets and
Markov Models...

Oh my!

@joshmeyerphd



The dog



th uh d ah g



The dog

FEATURE EXTRACTION

GMM MONOPHONE TRAINING

GMM TRIPHONE TRAINING

DNN TRAINING

DECODING

FEATURE EXTRACTION

- Sliding window feature extraction

GMM MONOPHONE TRAINING

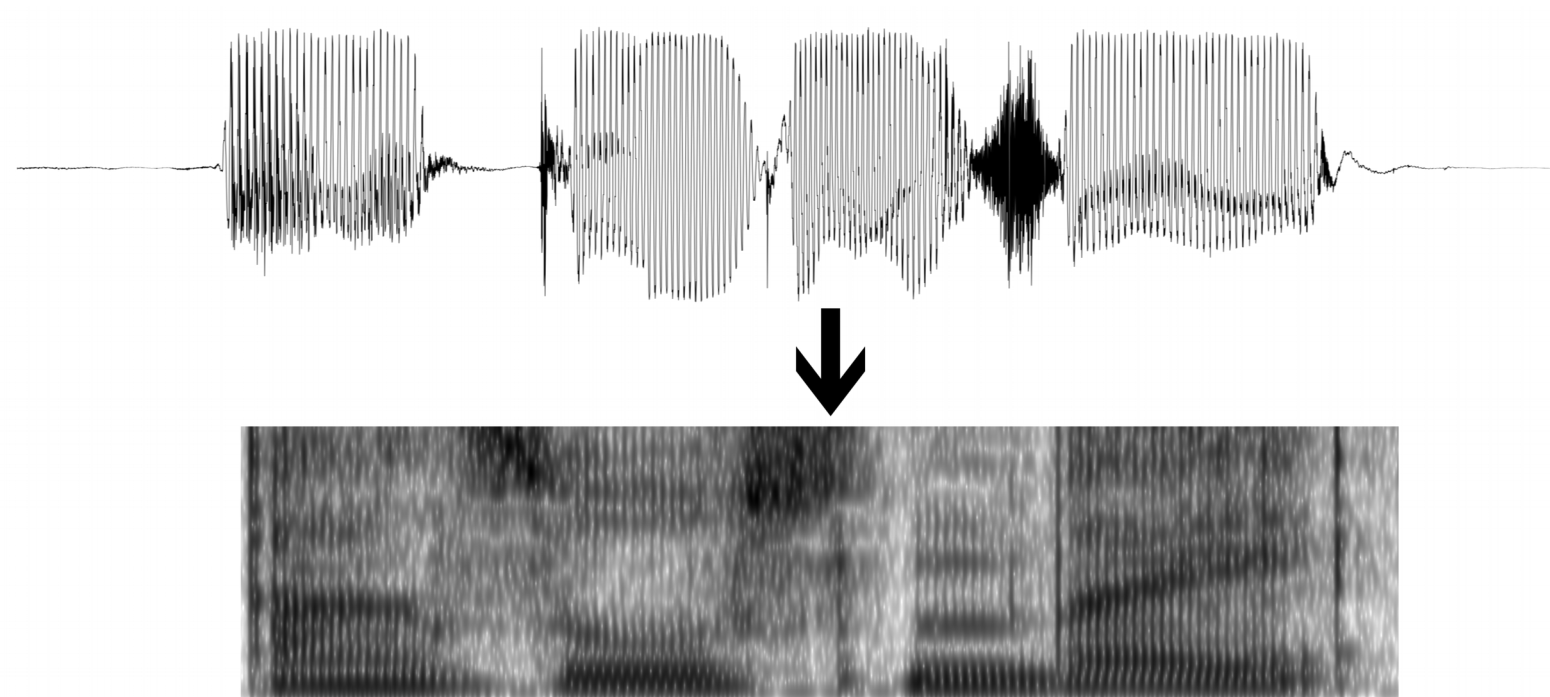
GMM TRIPHONE TRAINING

DNN TRAINING

DECODING

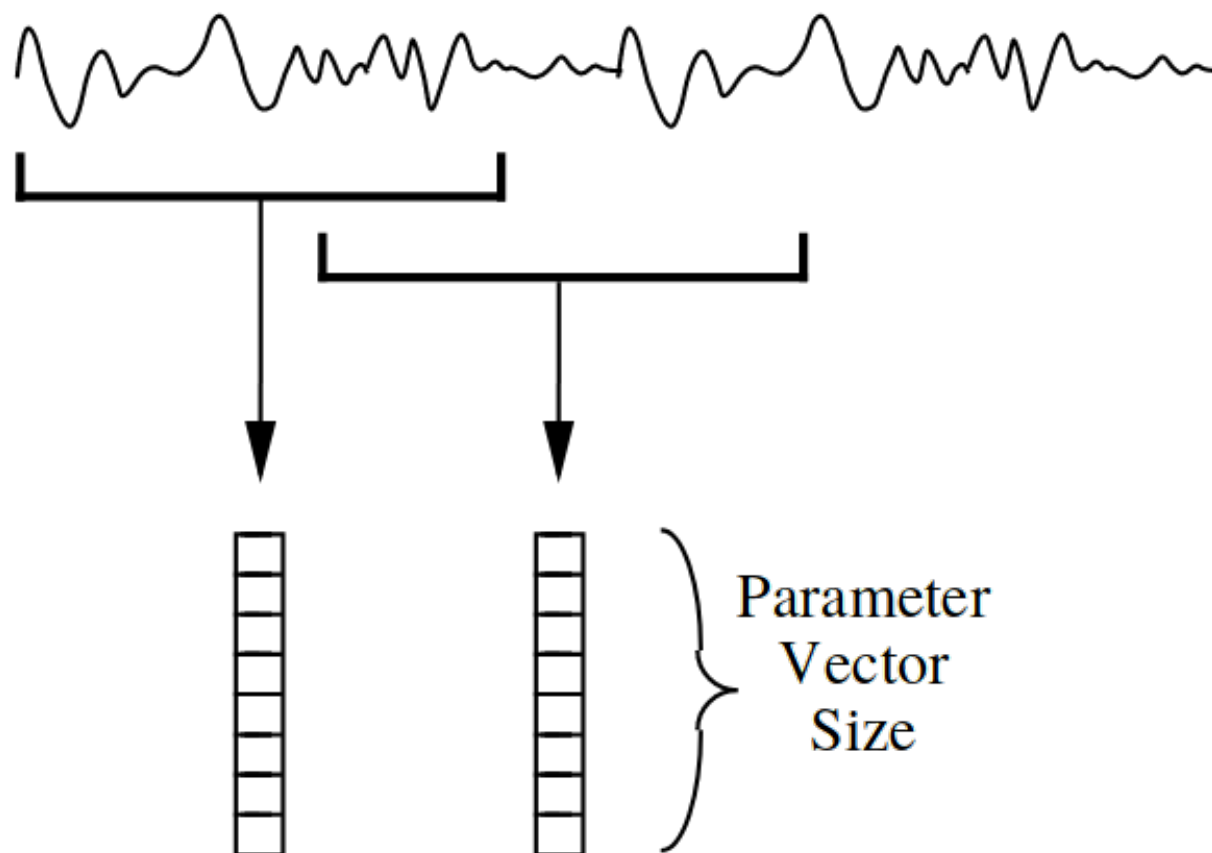
FEATURE EXTRACTION

- Sliding window feature extraction



FEATURE EXTRACTION

- Sliding window feature extraction



FEATURE EXTRACTION

- Sliding window feature extraction

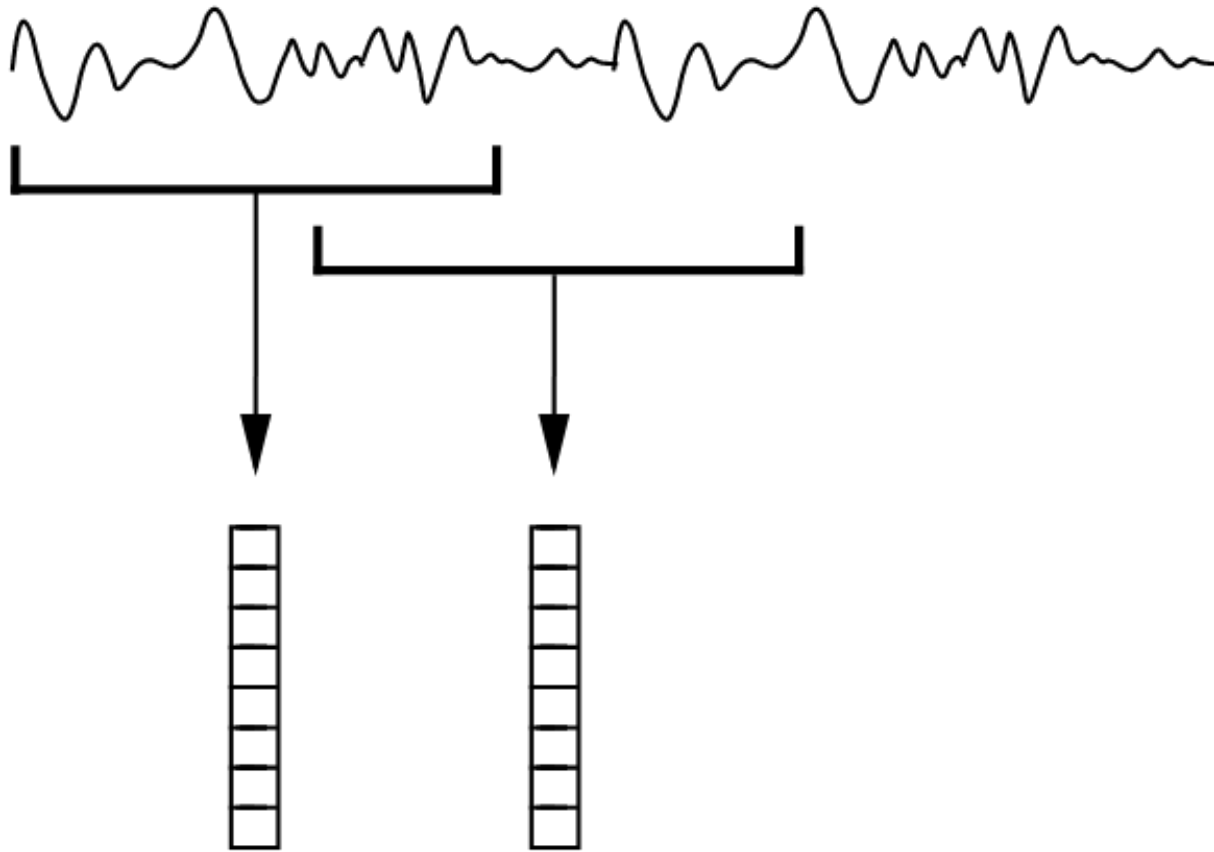
GMM MONOPHONE TRAINING

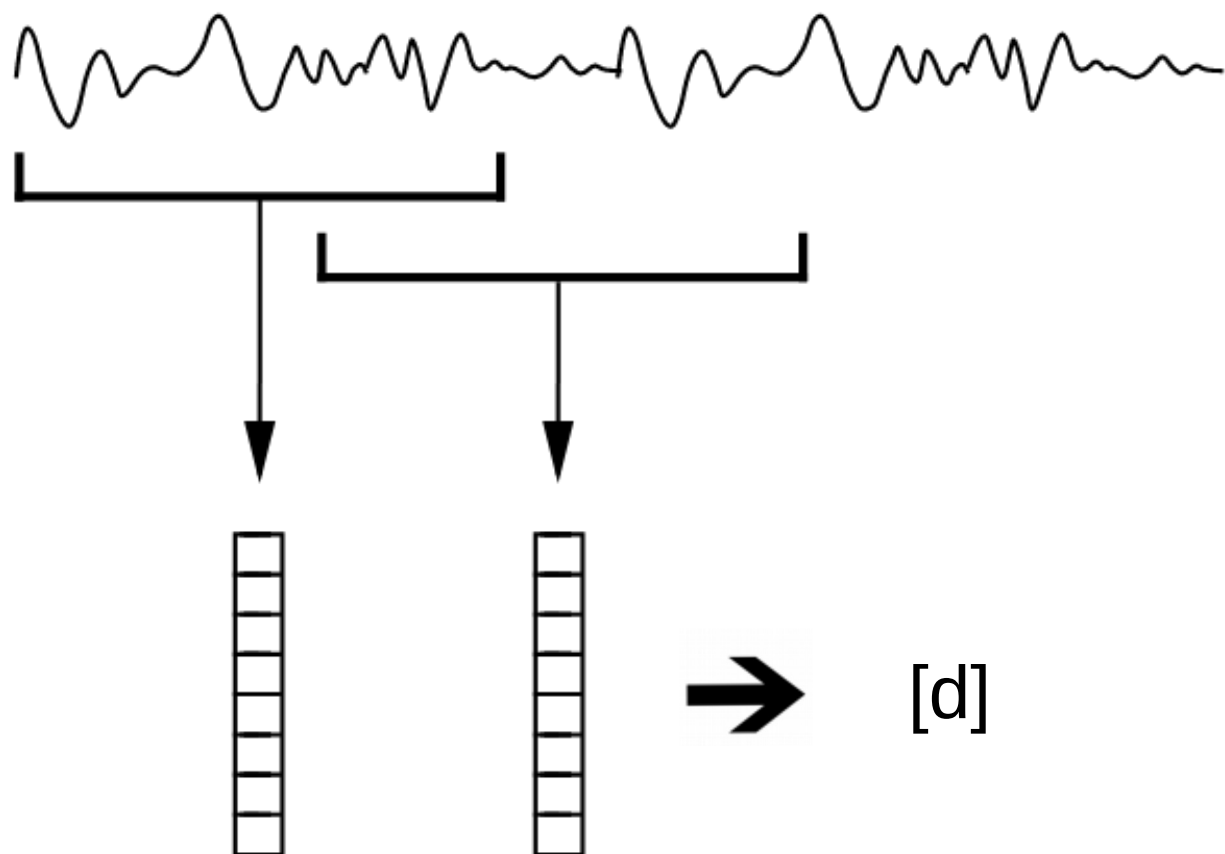
- Flat-start
- Baum-Welch re-estimation

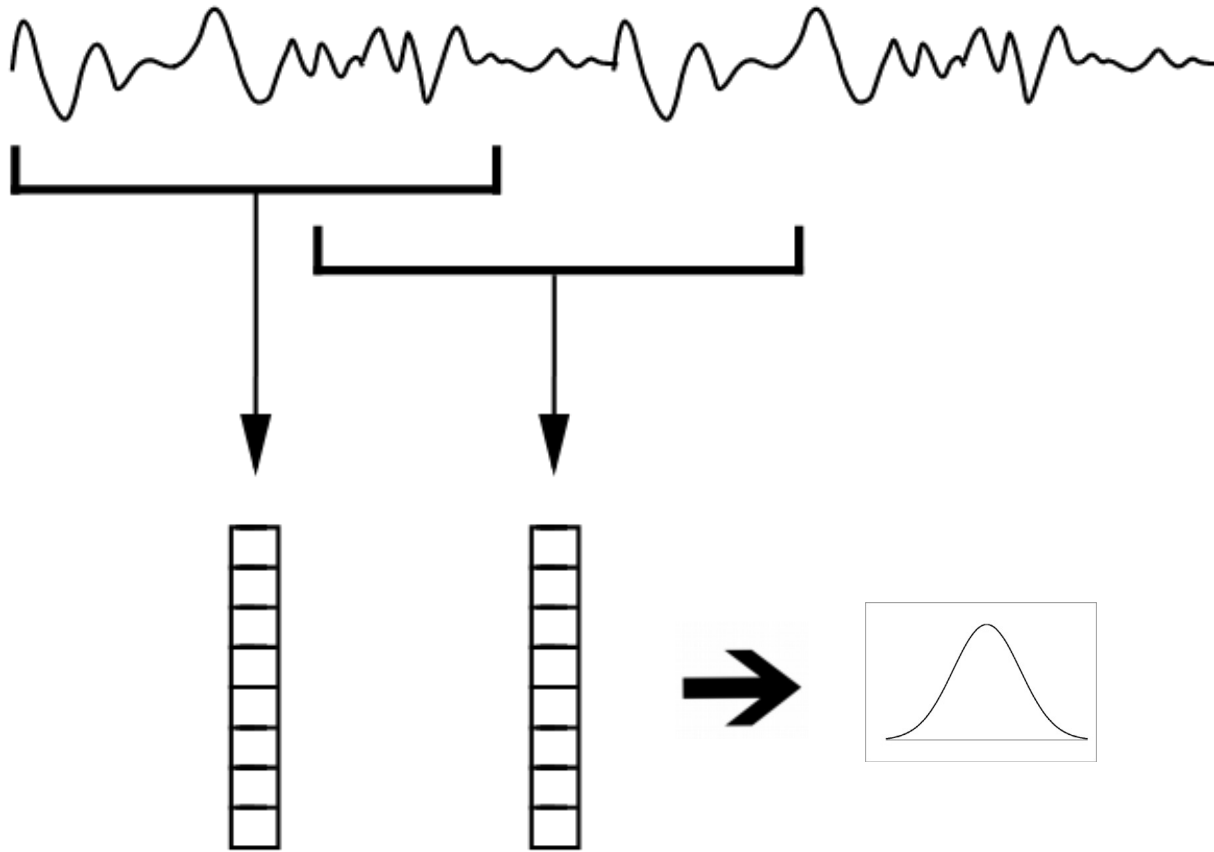
GMM TRIPHONE TRAINING

DNN TRAINING

DECODING







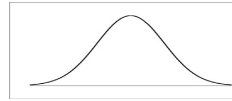
[t]

==



[p]

==



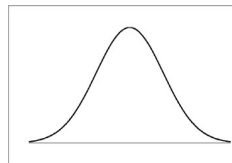
[k]

==



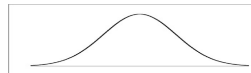
[a]

==



[i]

==

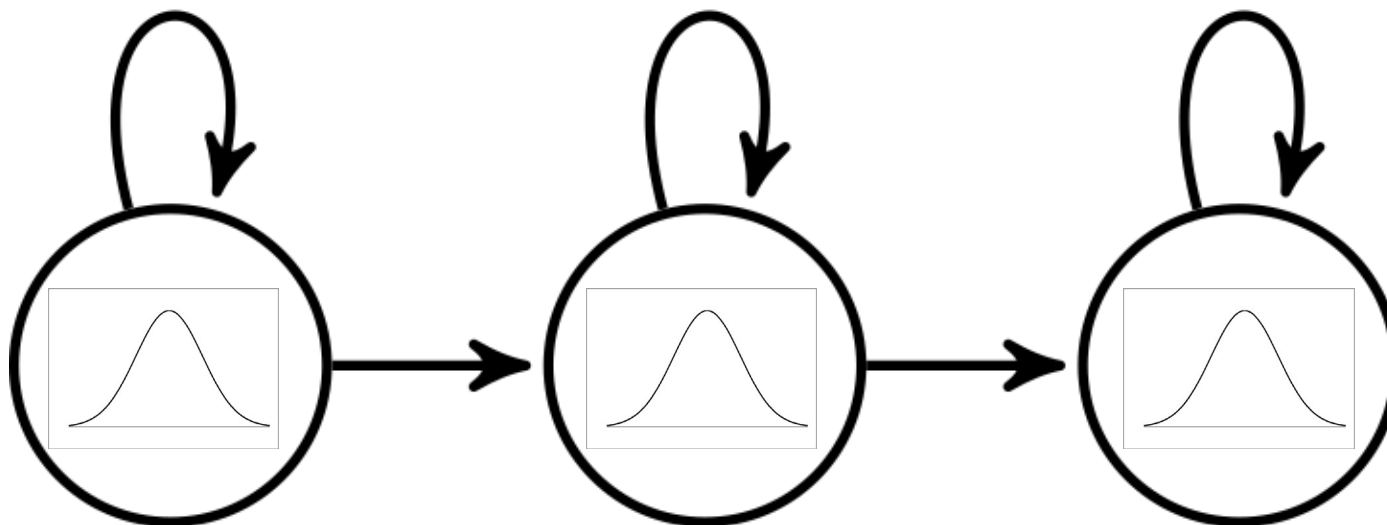


[u]

==

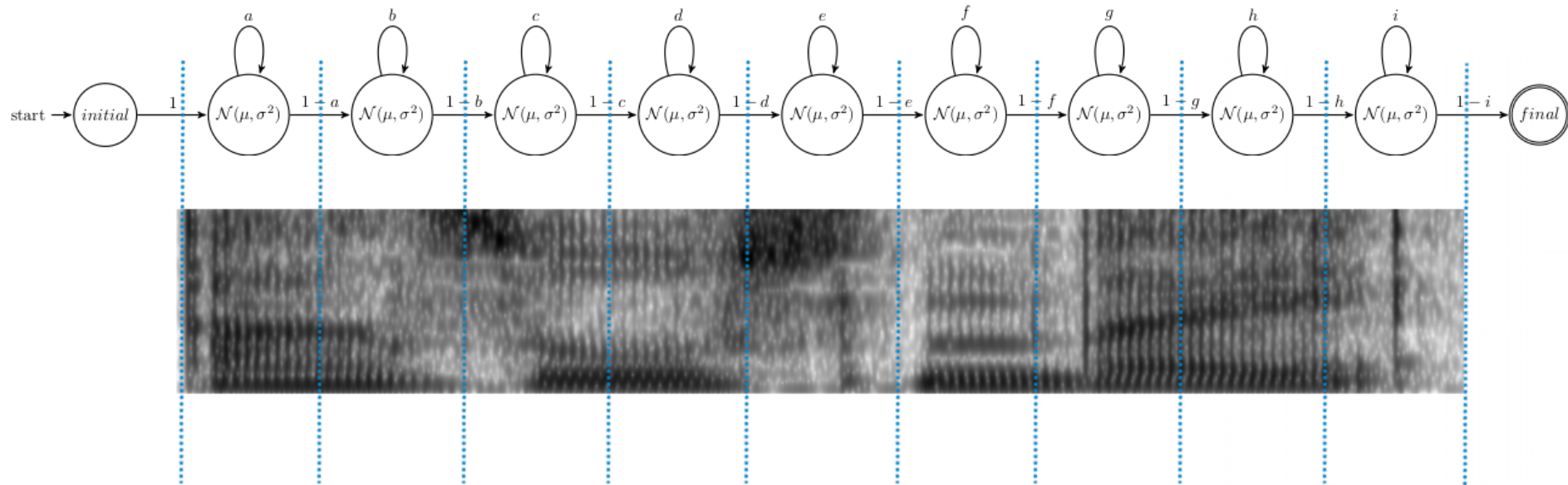


GMM MONOPHONE TRAINING



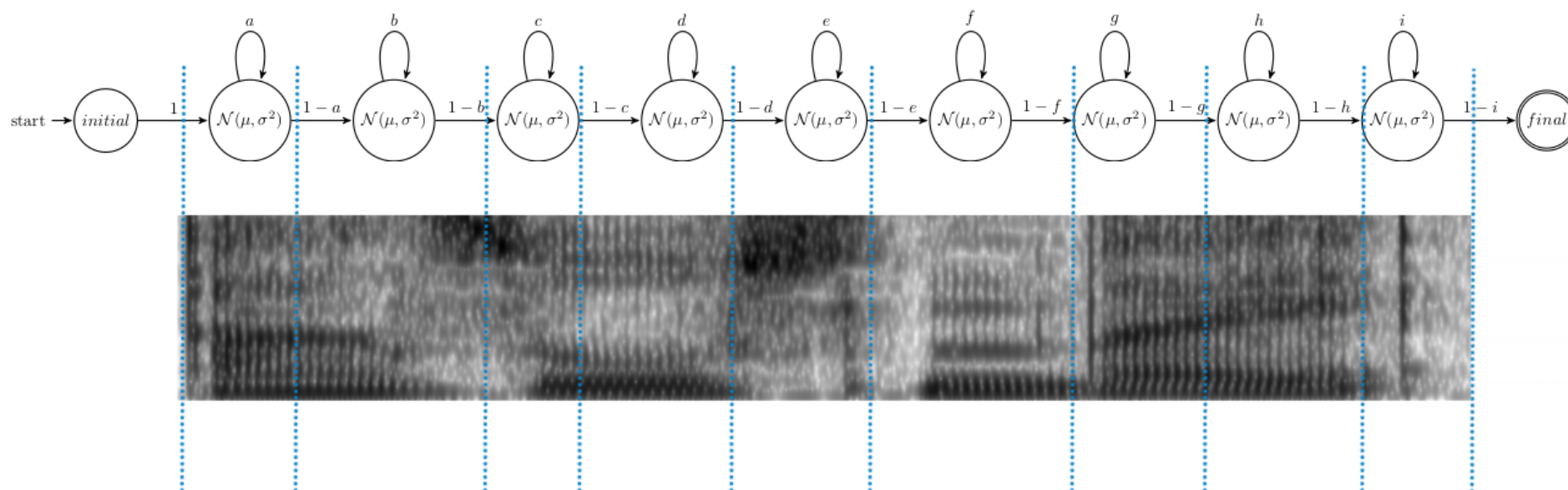
GMM MONOPHONE TRAINING

- Flat-start



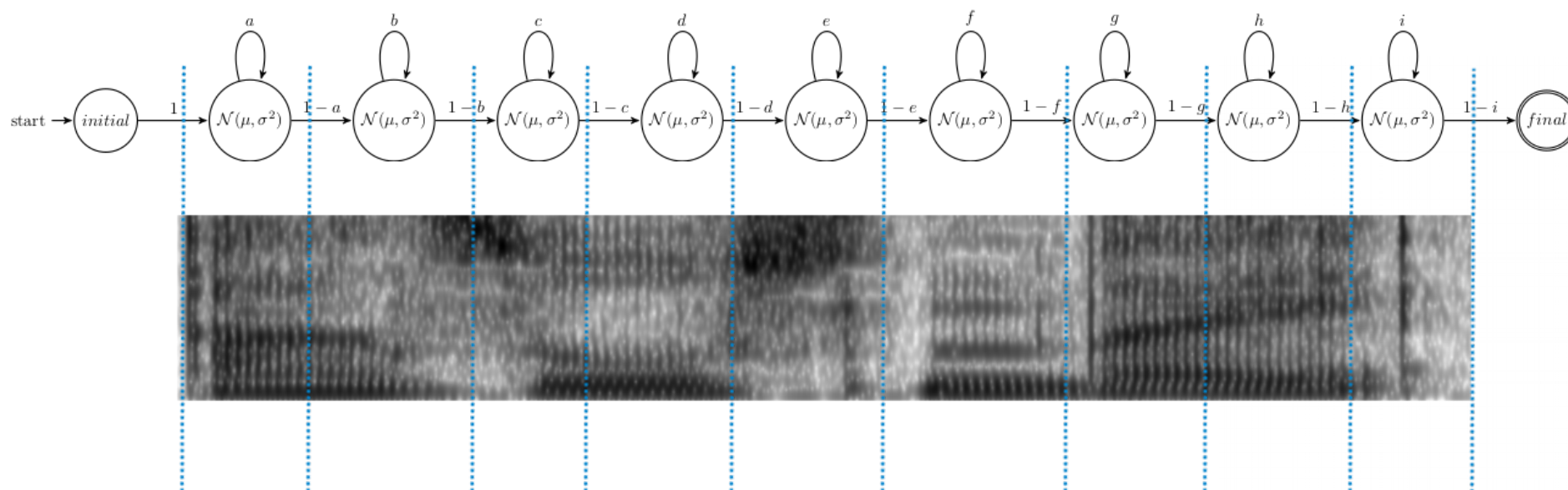
GMM MONOPHONE TRAINING

- Flat-start
- Baum-Welch Re-estimation



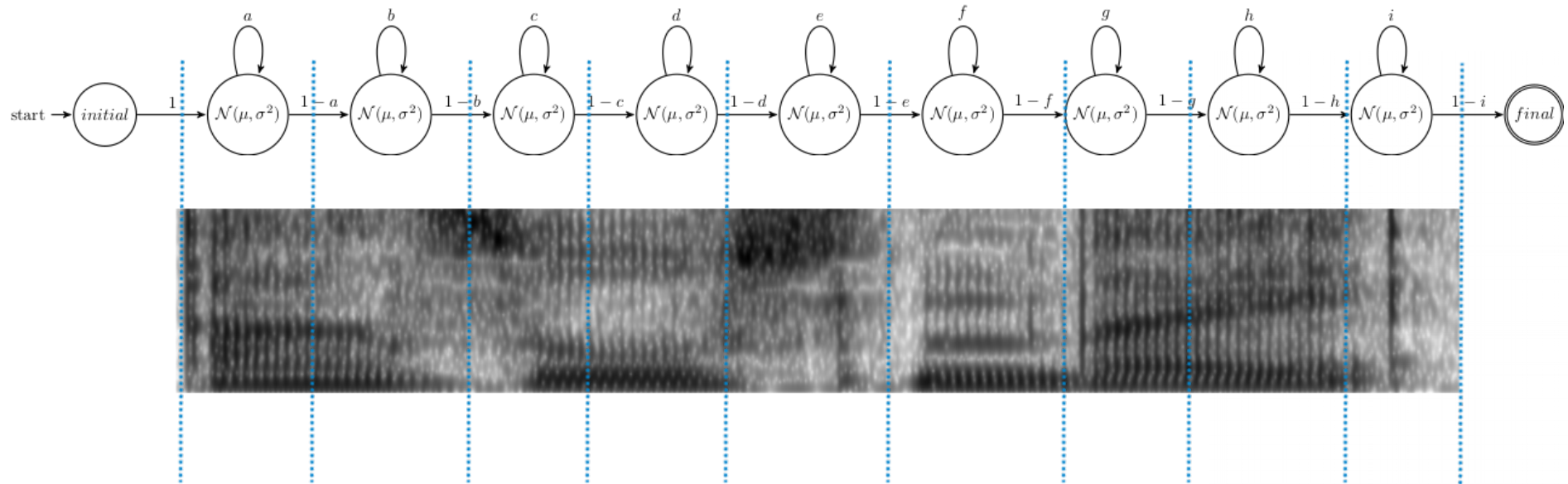
GMM MONOPHONE TRAINING

- Flat-start
- Baum-Welch Re-estimation



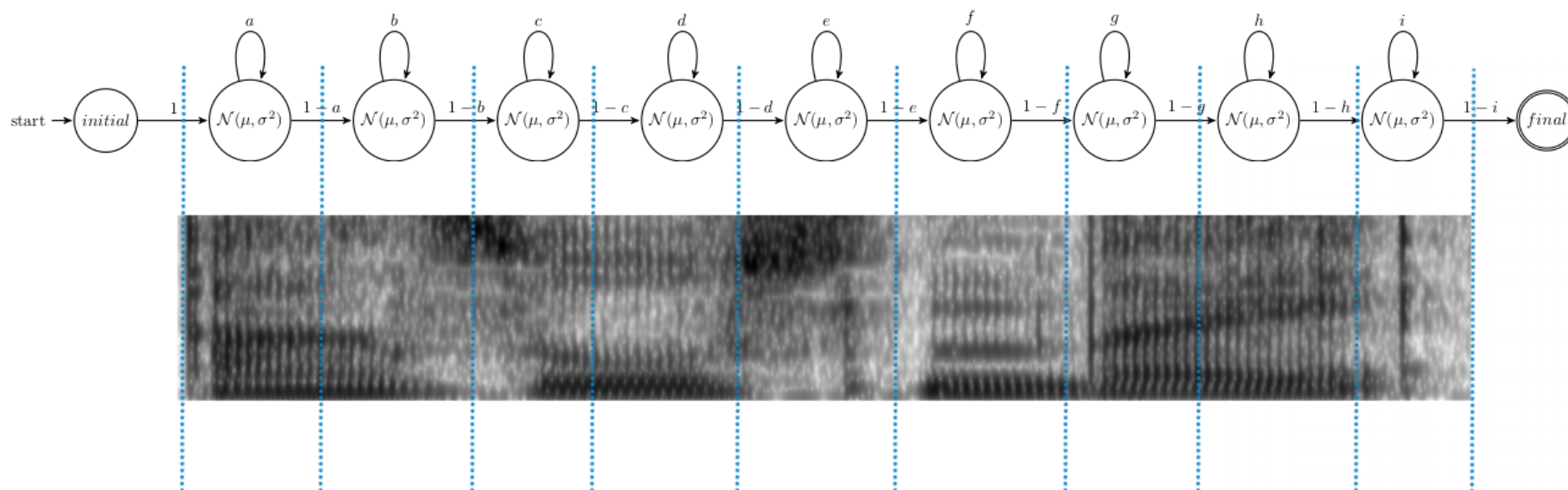
GMM MONOPHONE TRAINING

- Flat-start
- Baum-Welch Re-estimation



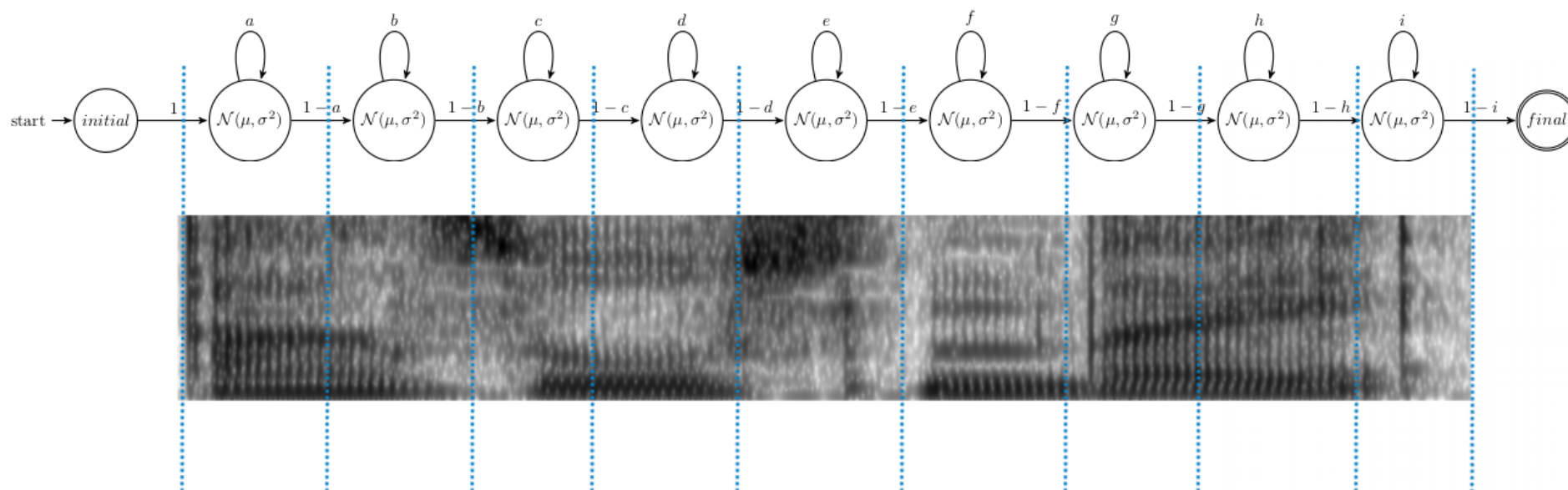
GMM MONOPHONE TRAINING

- Flat-start
- Baum-Welch Re-estimation



GMM MONOPHONE TRAINING

- Flat-start
- Baum-Welch Re-estimation



FEATURE EXTRACTION

- Sliding window feature extraction

GMM MONOPHONE TRAINING

- Flat-start
- Baum-Welch re-estimation

GMM TRIPHONE TRAINING

- Phonetic decision tree
- Baum-Welch re-estimation

DNN TRAINING

DECODING

GMM TRIPHONE TRAINING

- Phonetic decision tree

* -aw - *

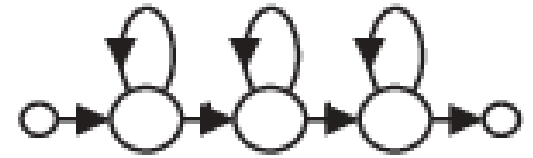


Figure credit:
The HTK Book

GMM TRIPHONE TRAINING

- Phonetic decision tree

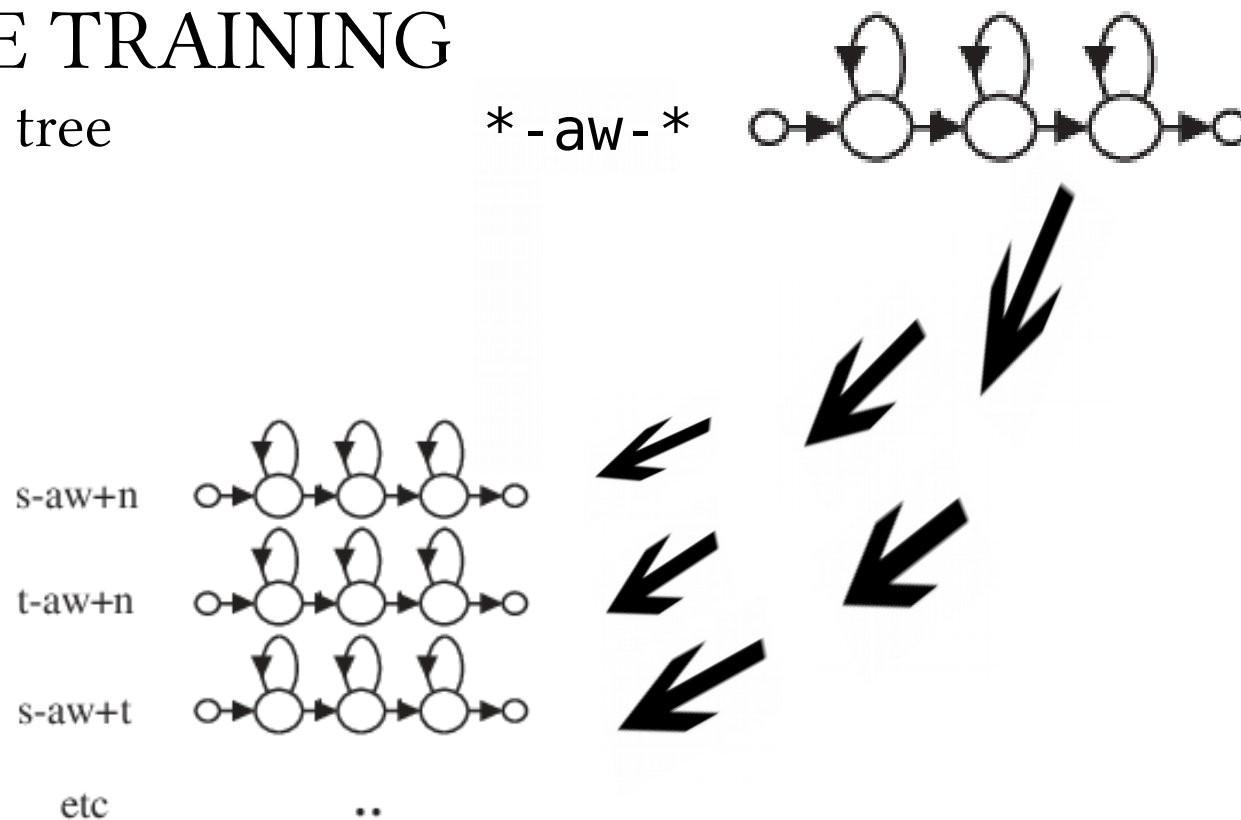


Figure credit:
The HTK Book

GMM TRIPHONE TRAINING

- Phonetic decision tree

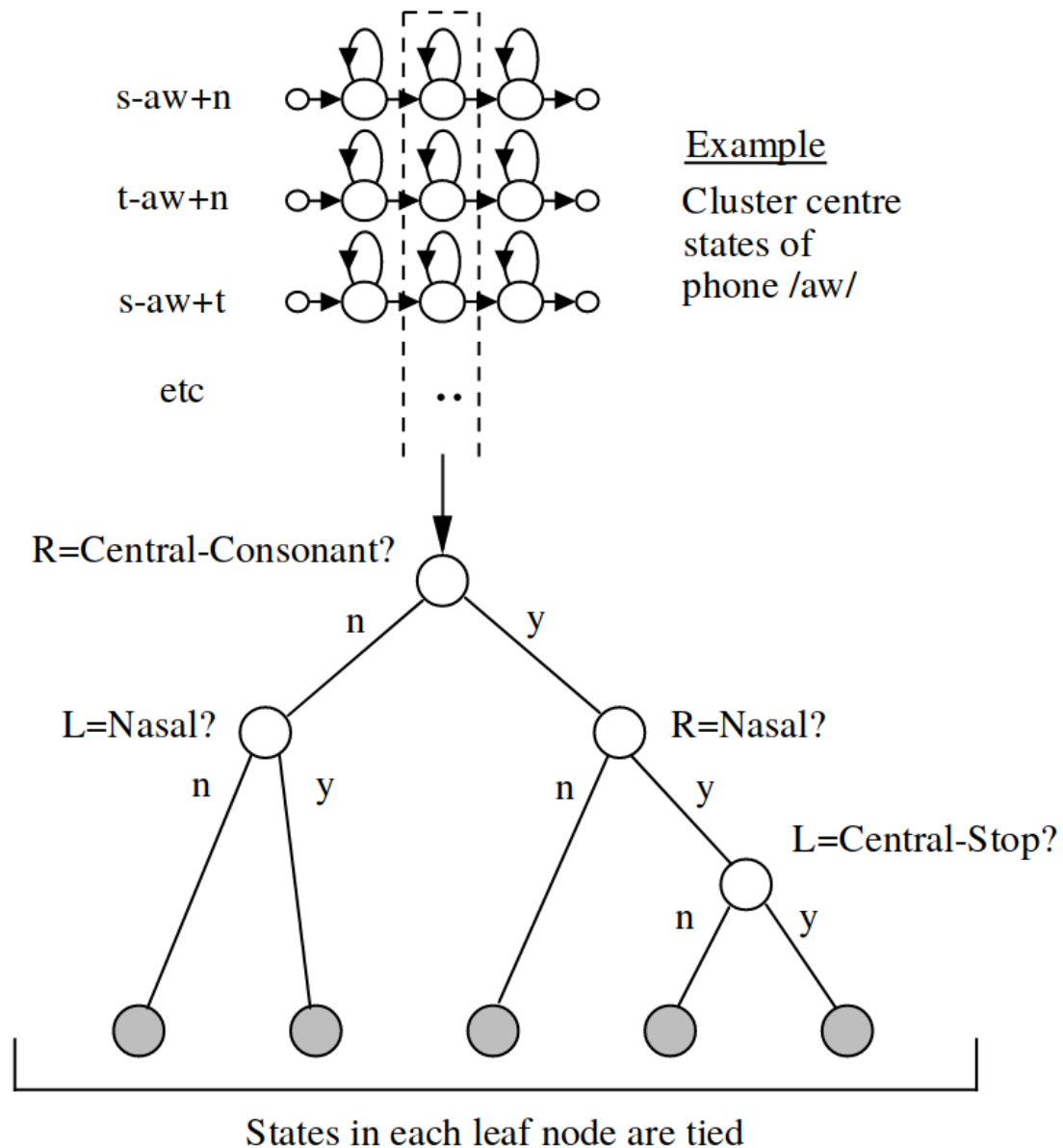
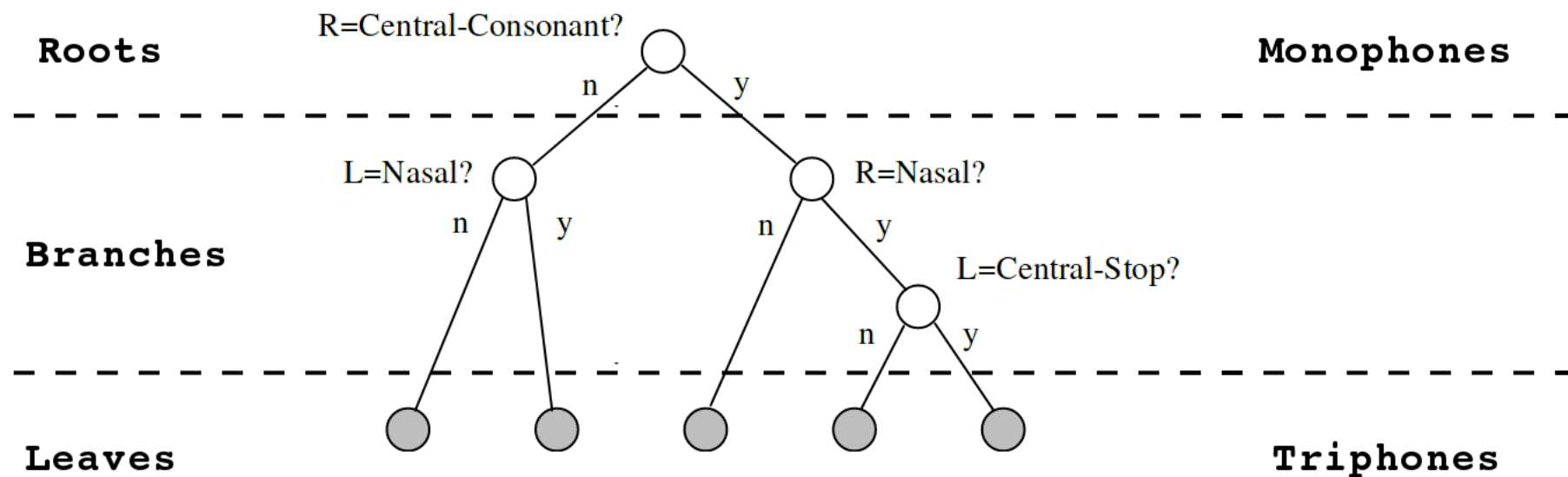


Figure credit:
The HTK Book

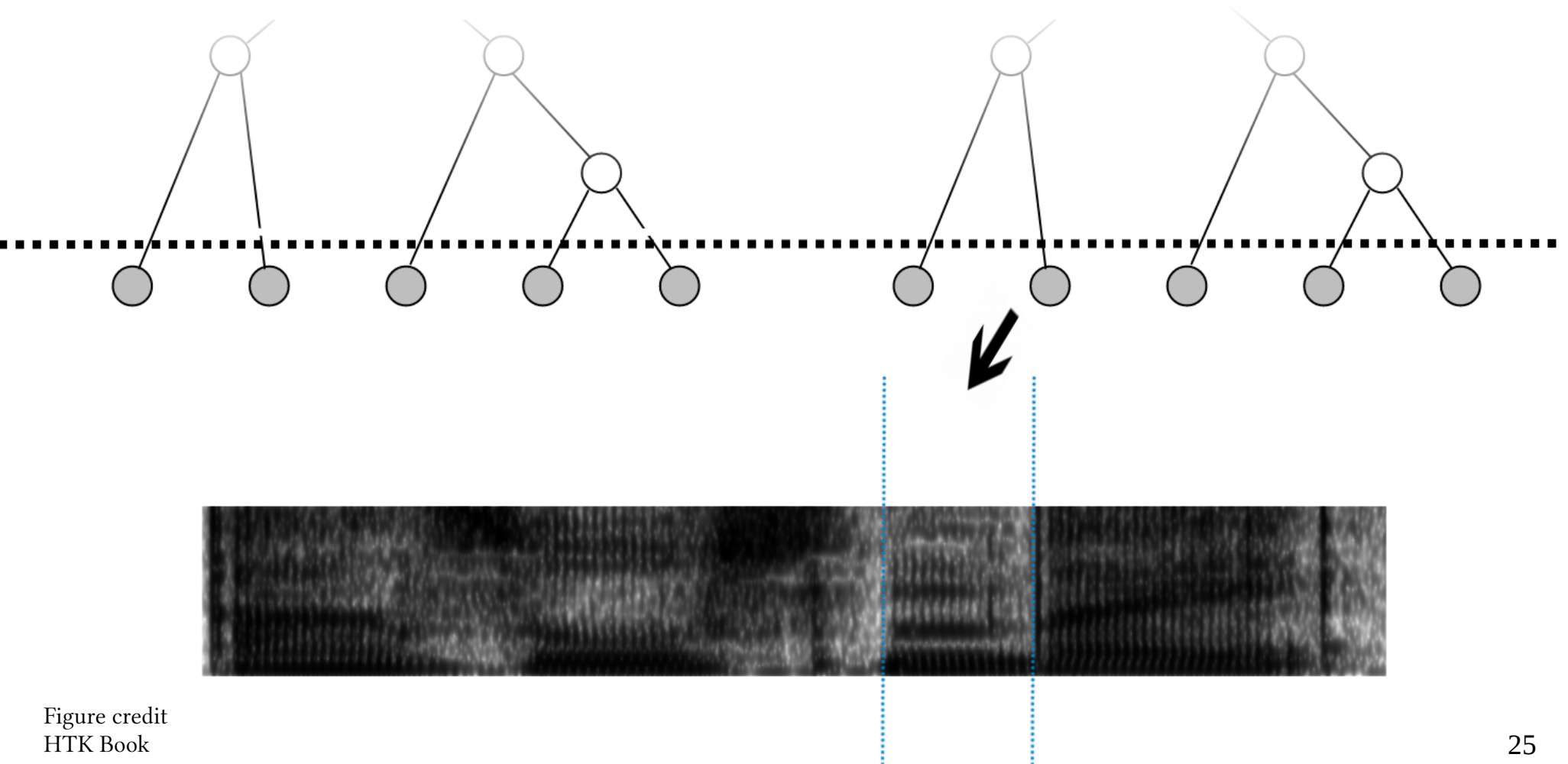
GMM TRIPHONE TRAINING

- Phonetic decision tree



GMM TRIPHONE TRAINING

- Phonetic decision tree
- Baum-Welch re-estimation



FEATURE EXTRACTION

- Sliding window feature extraction

GMM MONOPHONE TRAINING

- Flat-start
- Baum-Welch re-estimation

GMM TRIPHONE TRAINING

- Phonetic decision tree
- Baum-Welch re-estimation

DNN TRAINING

- Labeled Data
- Gradient descent via backprop

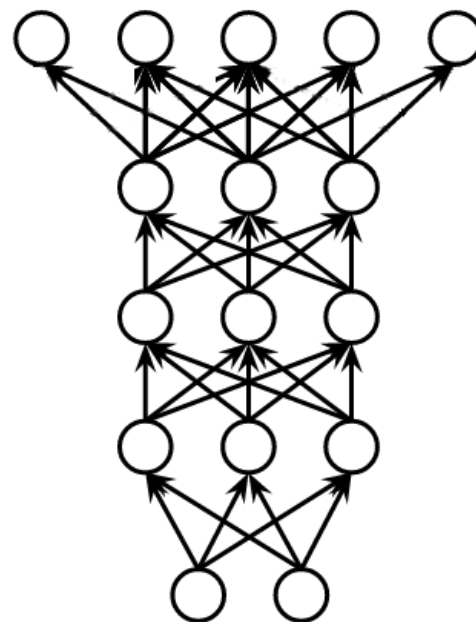
DECODING

DNN TRAINING

- Labeled Data

Phonetic Labels

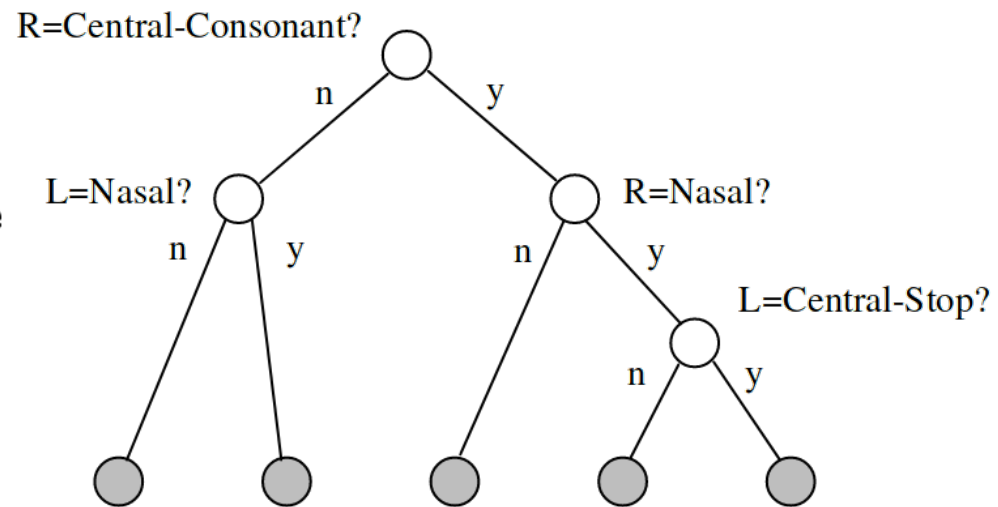
Audio Features



DNN TRAINING

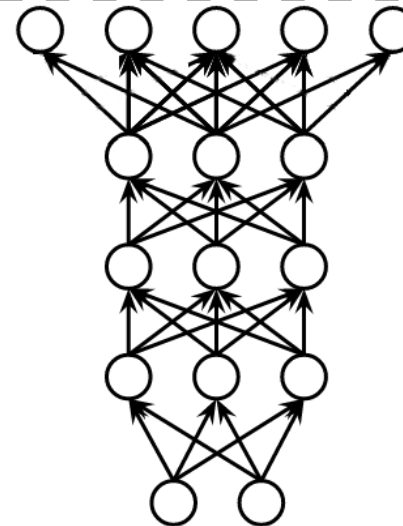
- Labeled Data

Phonetic Decision Tree



Label Equivalences

DNN Acoustic Model

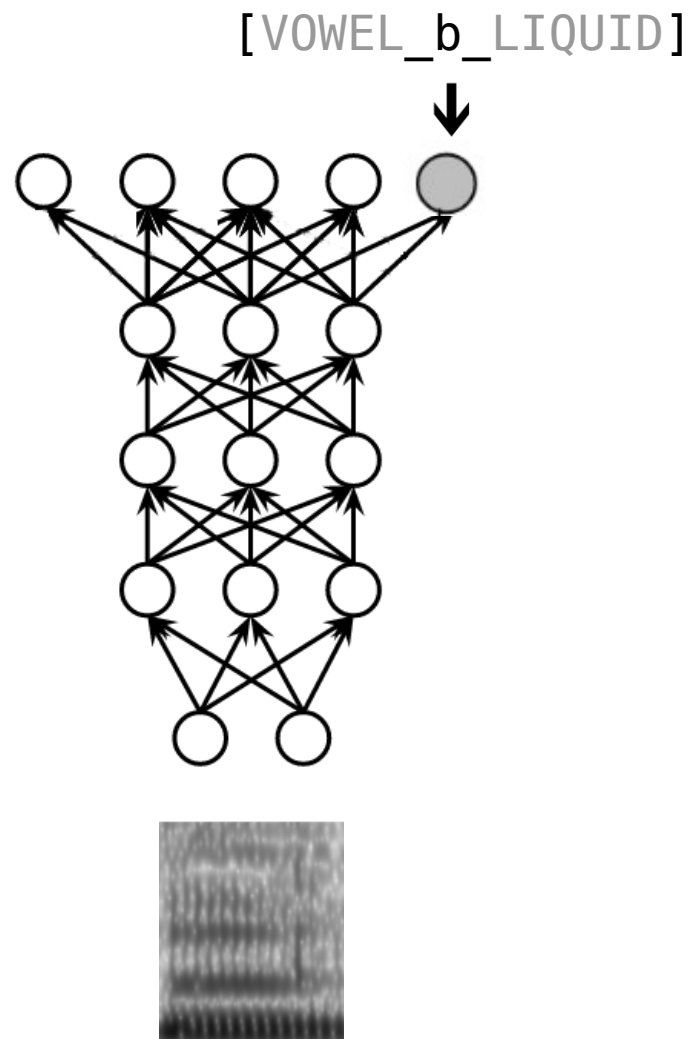


DNN TRAINING

- Labeled Data
- Gradient descent via backprop

Phonetic Labels

Audio Features



FEATURE EXTRACTION

- Sliding window feature extraction

GMM MONOPHONE TRAINING

- Flat-start
- Baum-Welch re-estimation

GMM TRIPHONE TRAINING

- Phonetic decision tree
- Baum-Welch re-estimation

DNN TRAINING

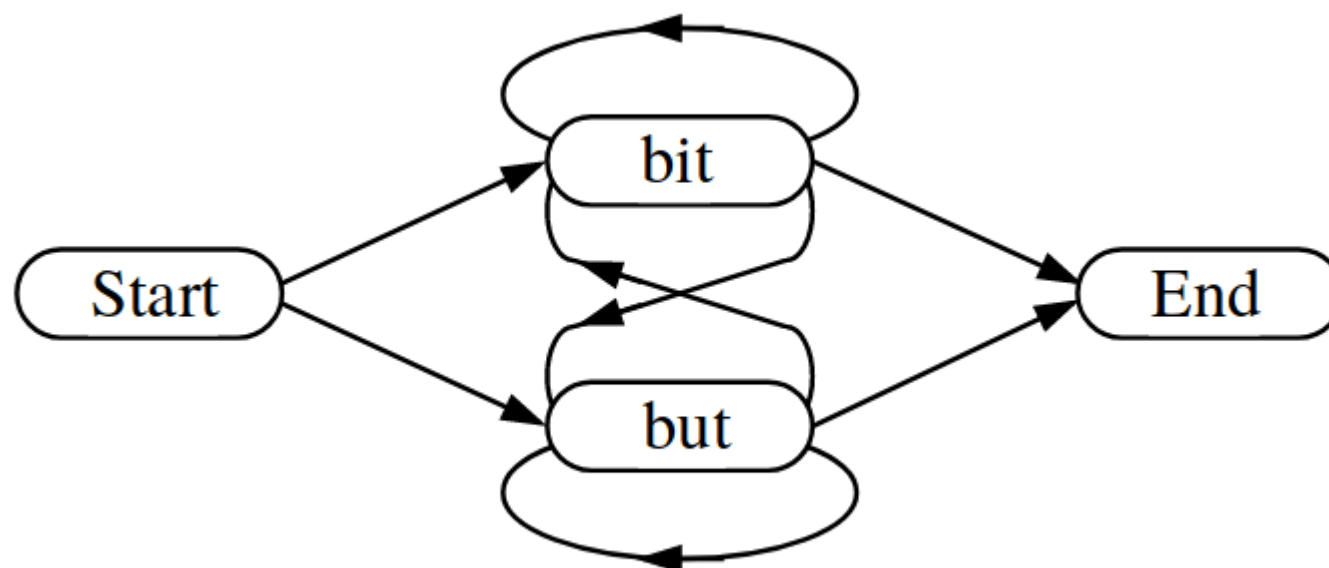
- Labeled Data
- Gradient descent via backprop

DECODING

- WFST Graph compilation
- Viterbi decoding

DECODING

- WFST Graph compilation



DECODING

- WFST Graph compilation

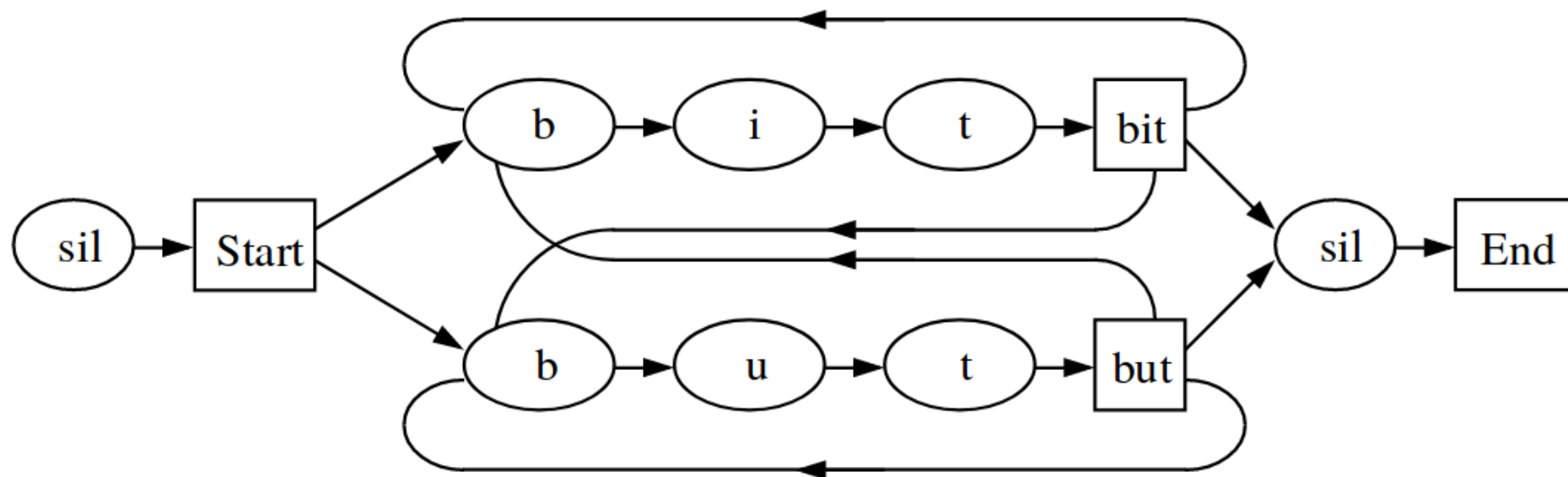


Figure credit
The HTK Book

DECODING

- WFST Graph compilation

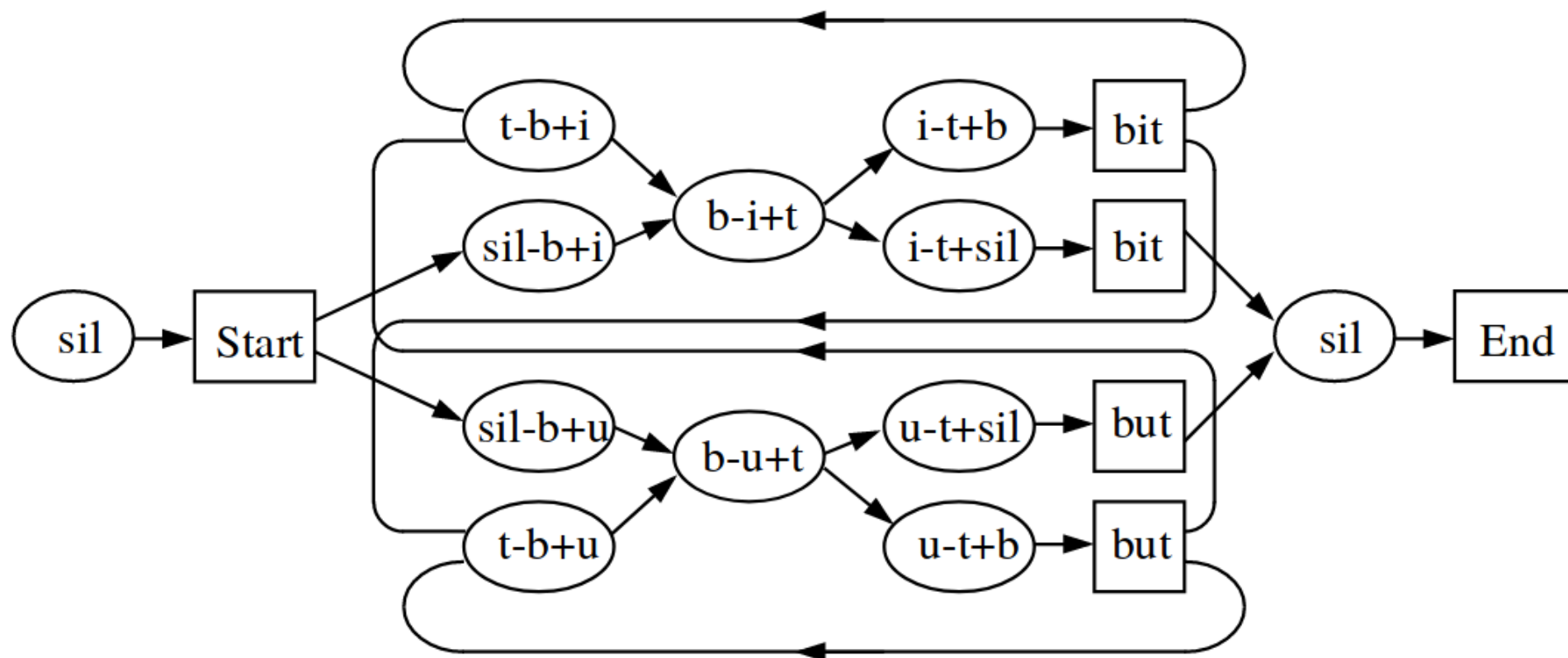


Figure credit
The HTK Book

DECODING

- WFST Graph compilation
- Viterbi decoding

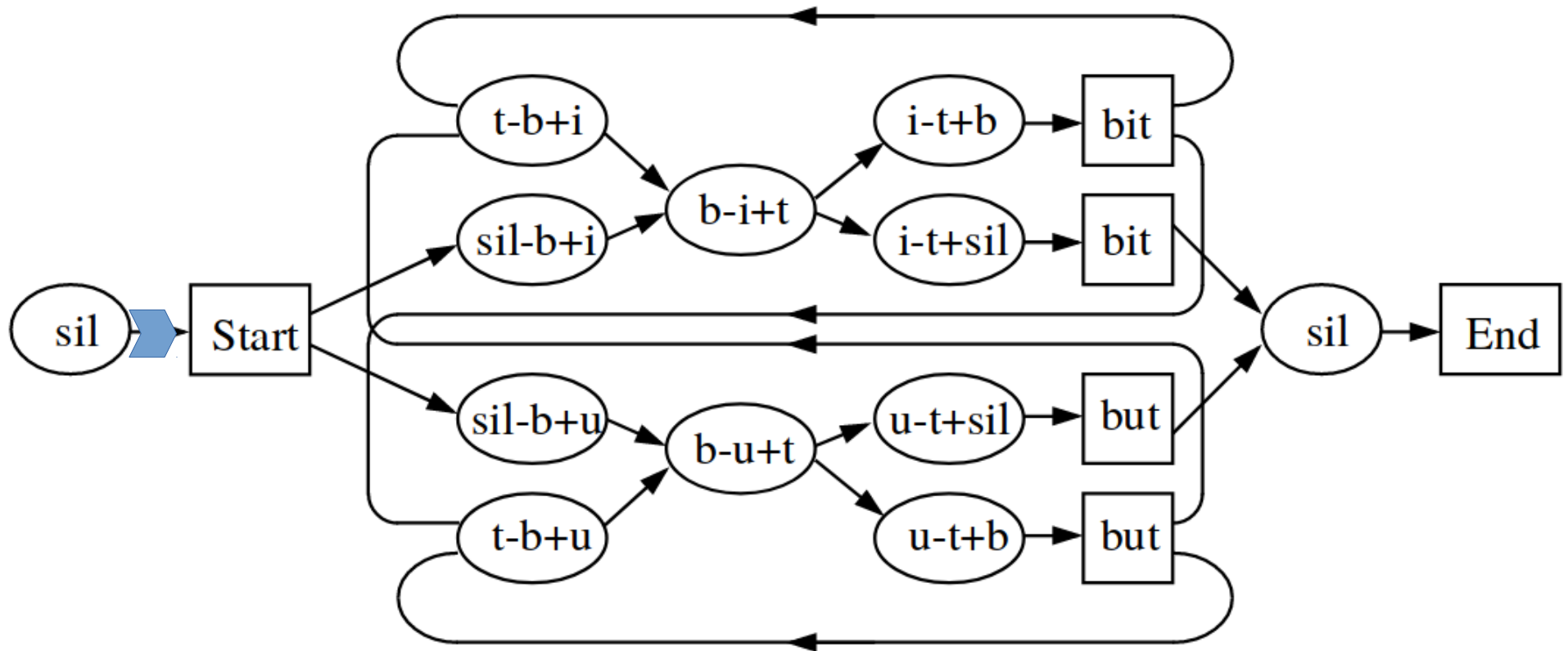


Figure credit
The HTK Book

DECODING

- WFST Graph compilation
- Viterbi decoding

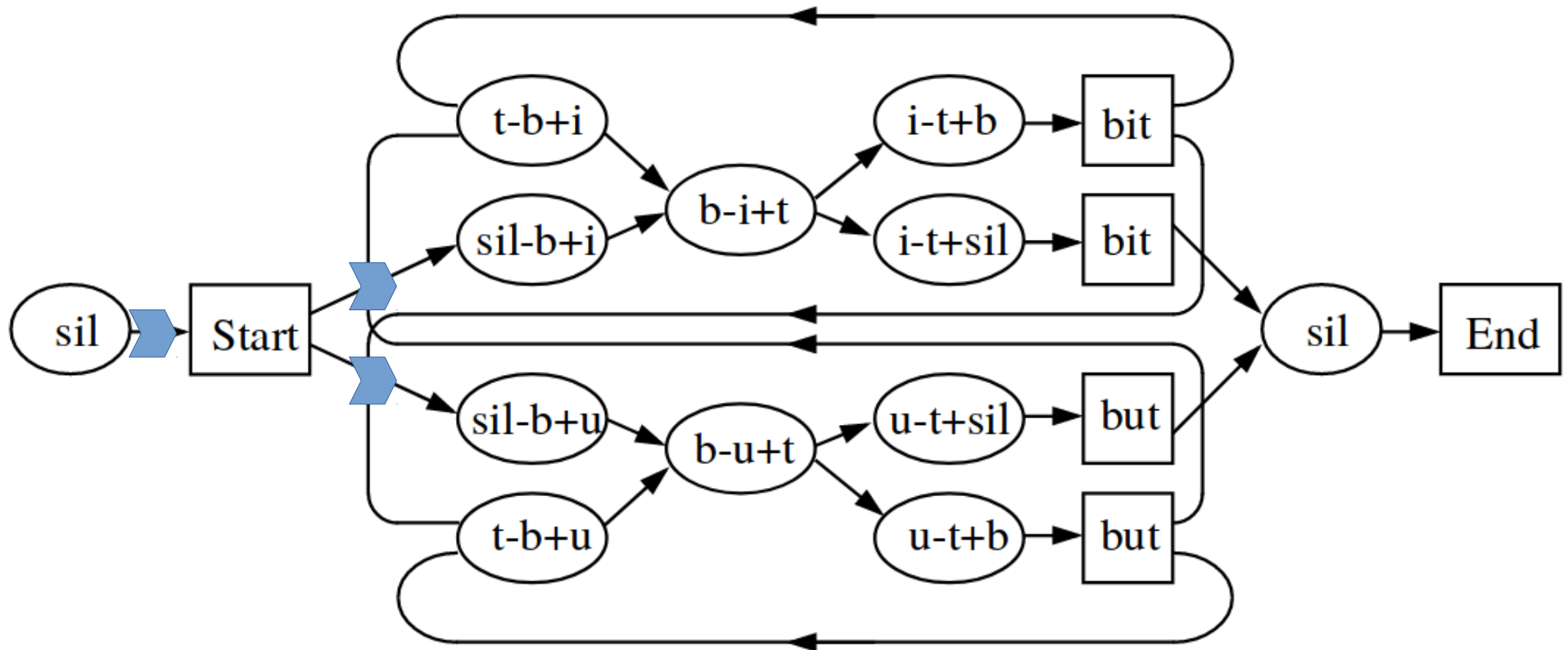


Figure credit
The HTK Book

DECODING

- WFST Graph compilation
- Viterbi decoding

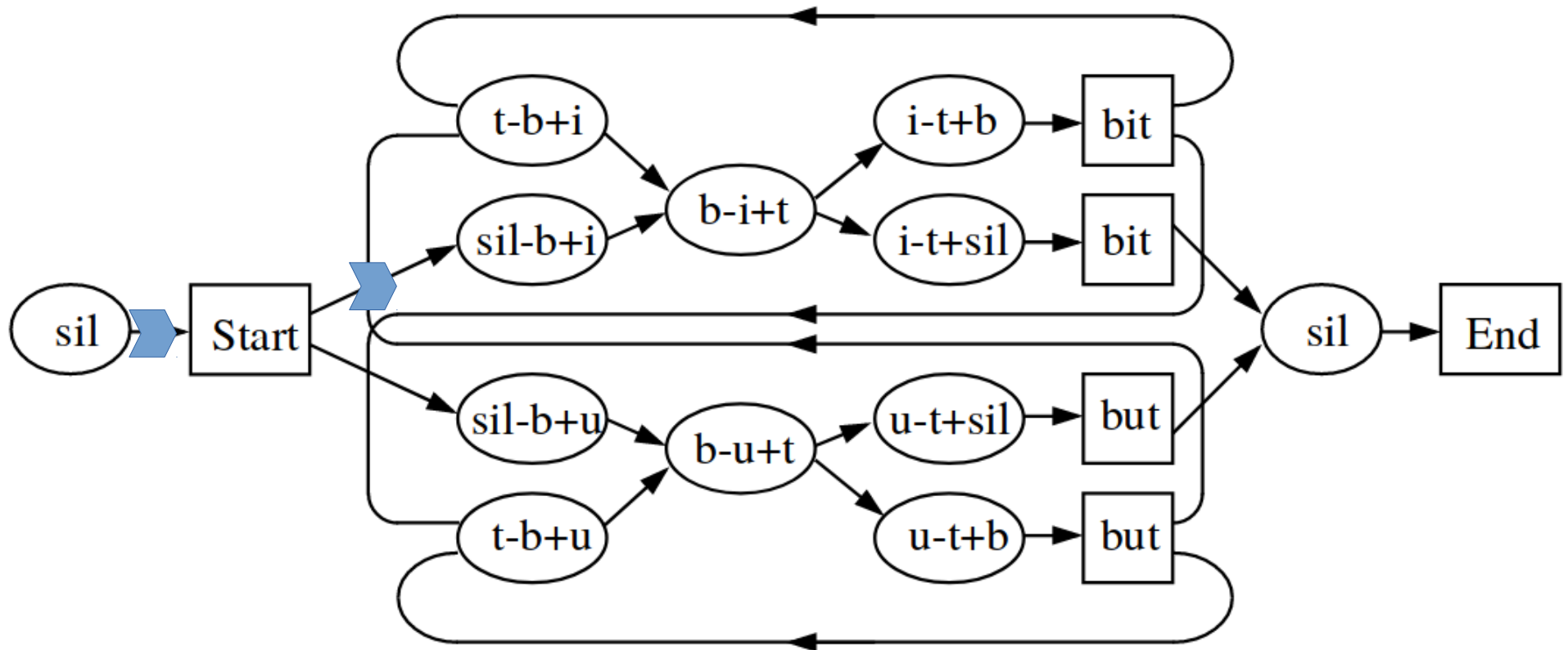


Figure credit
The HTK Book

DECODING

- WFST Graph compilation
- Viterbi decoding

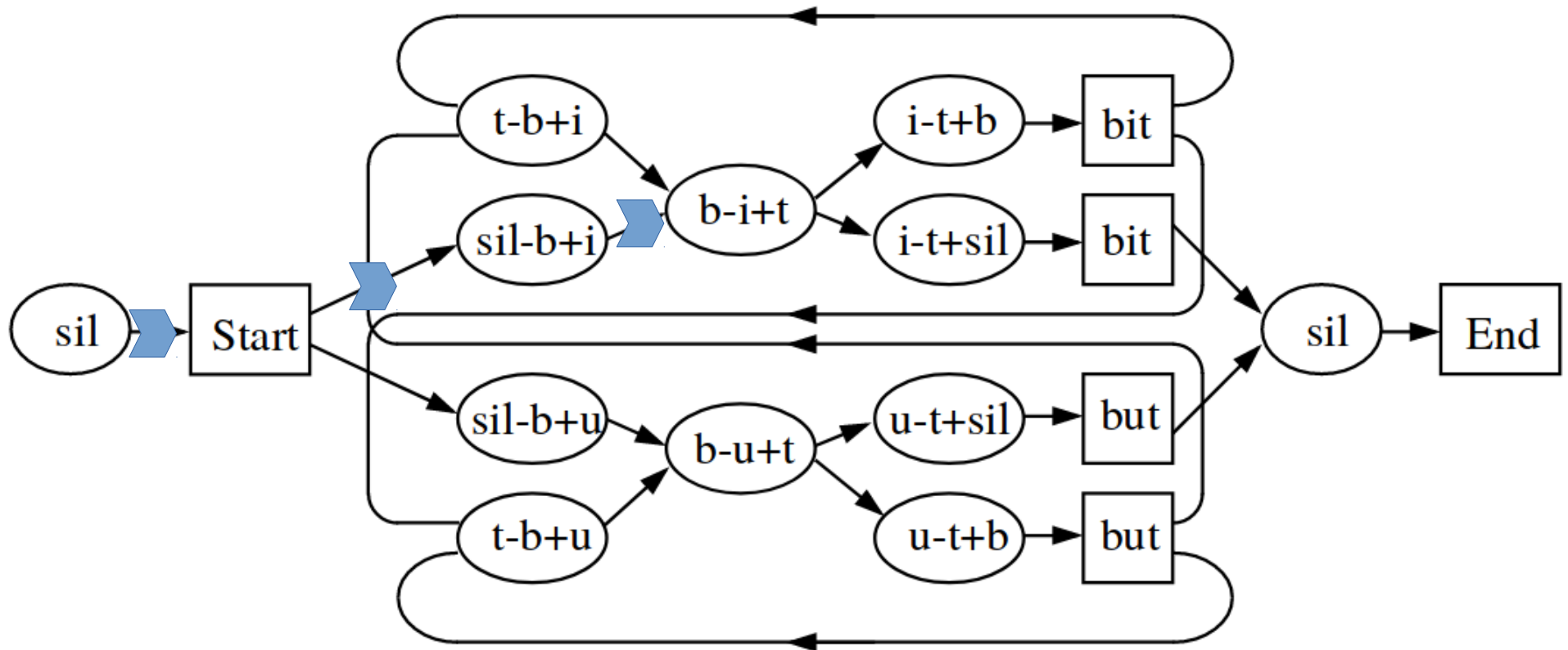


Figure credit
The HTK Book

DECODING

- WFST Graph compilation
- Viterbi decoding

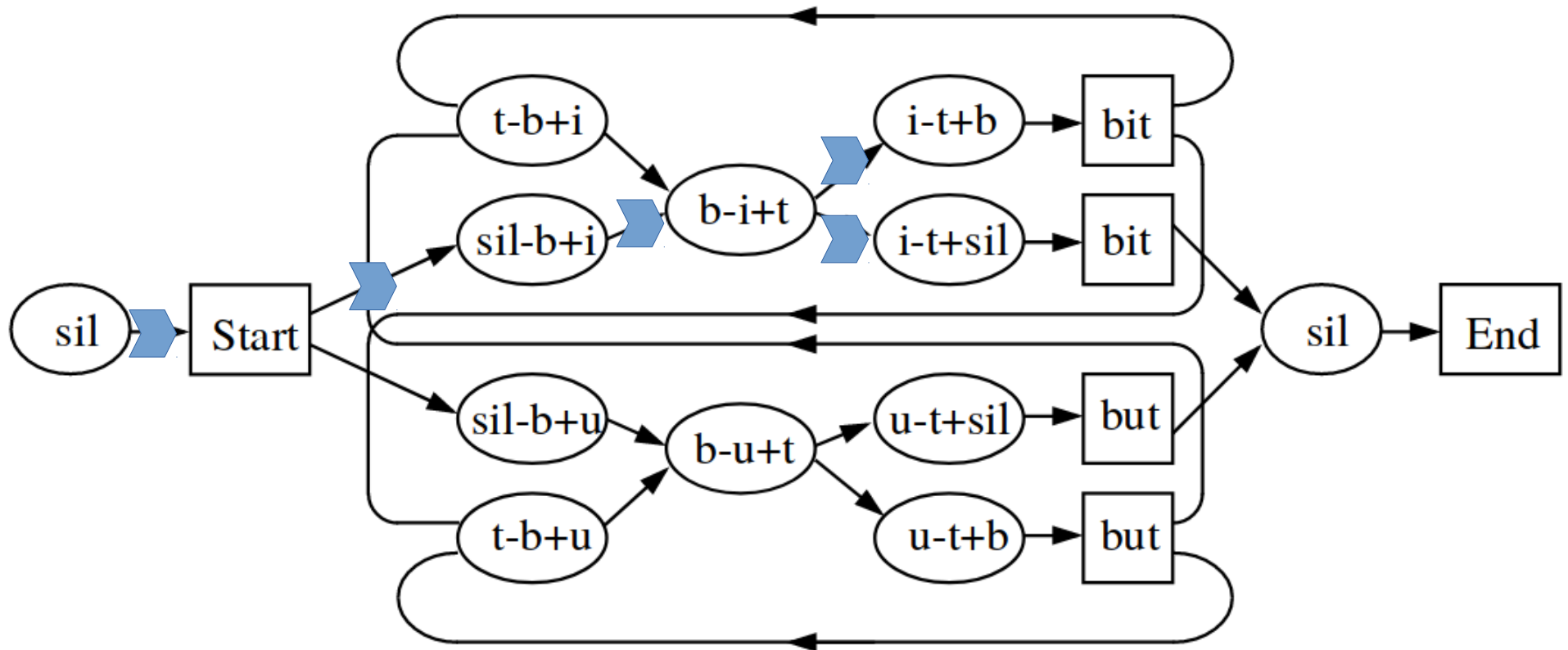


Figure credit
The HTK Book

DECODING

- WFST Graph compilation
- Viterbi decoding

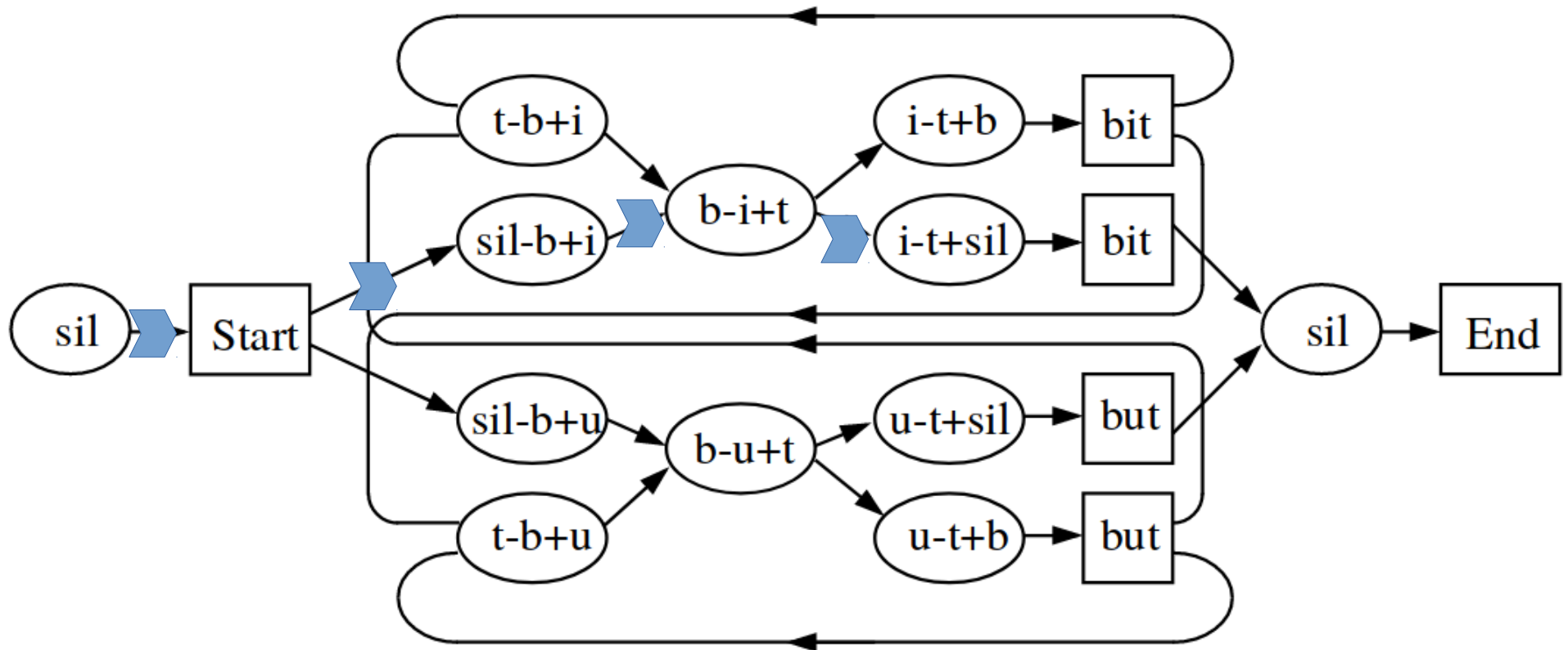


Figure credit
The HTK Book

DECODING

- WFST Graph compilation
- Viterbi decoding

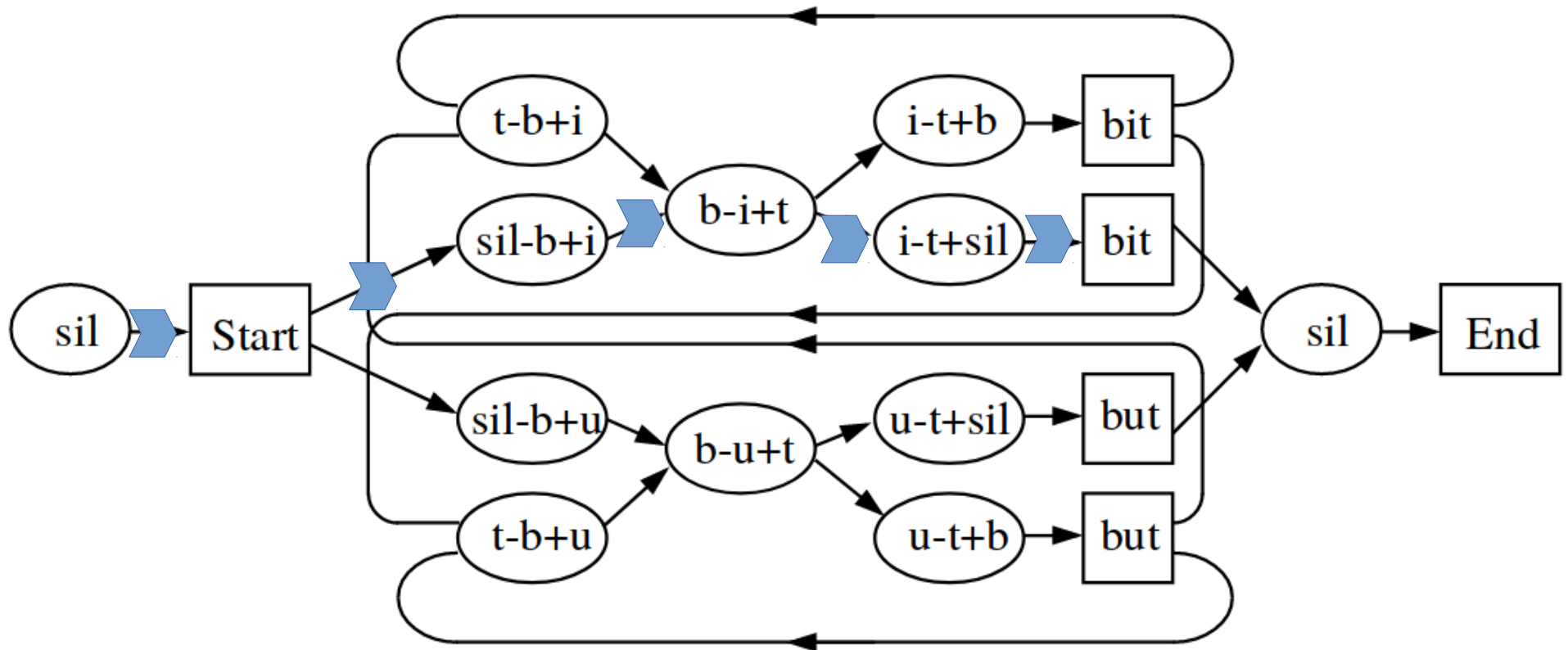


Figure credit
The HTK Book

DECODING

- WFST Graph compilation
- Viterbi decoding

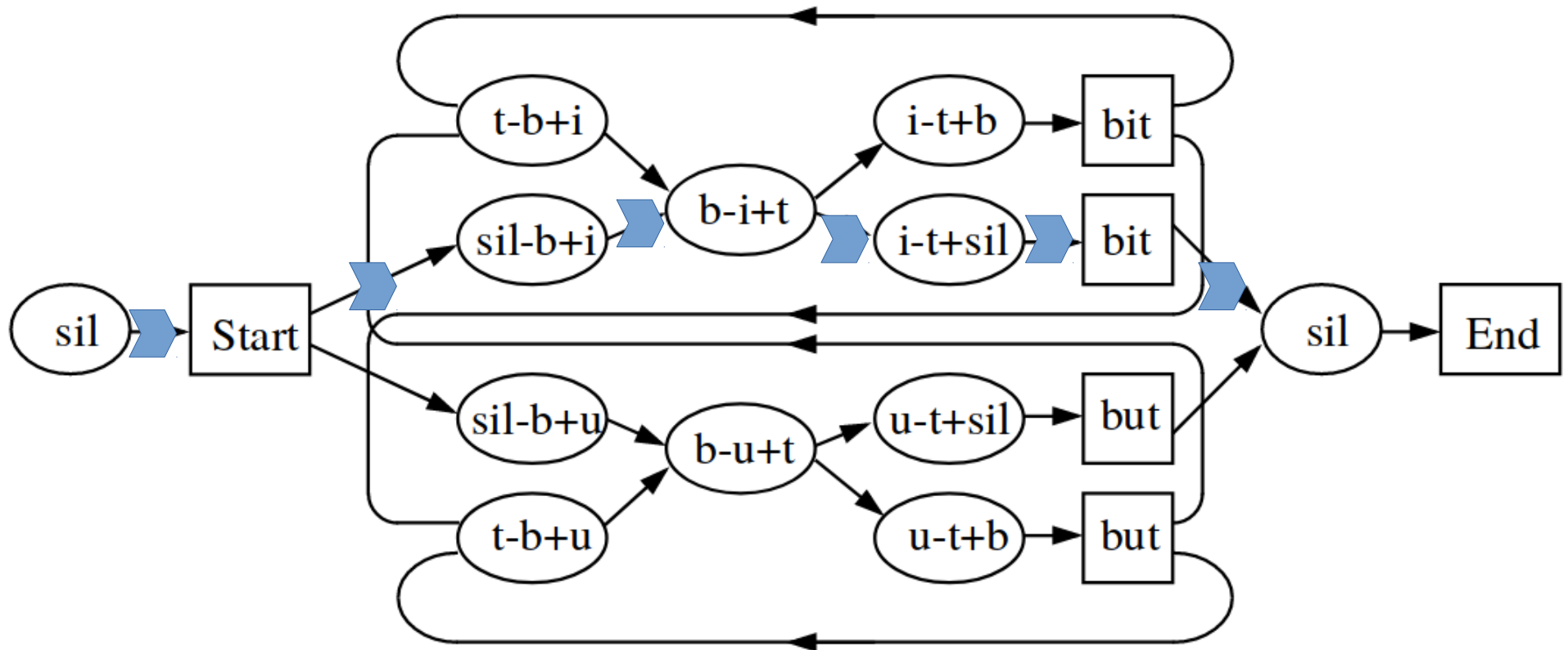


Figure credit
The HTK Book

DECODING

- WFST Graph compilation
- Viterbi decoding

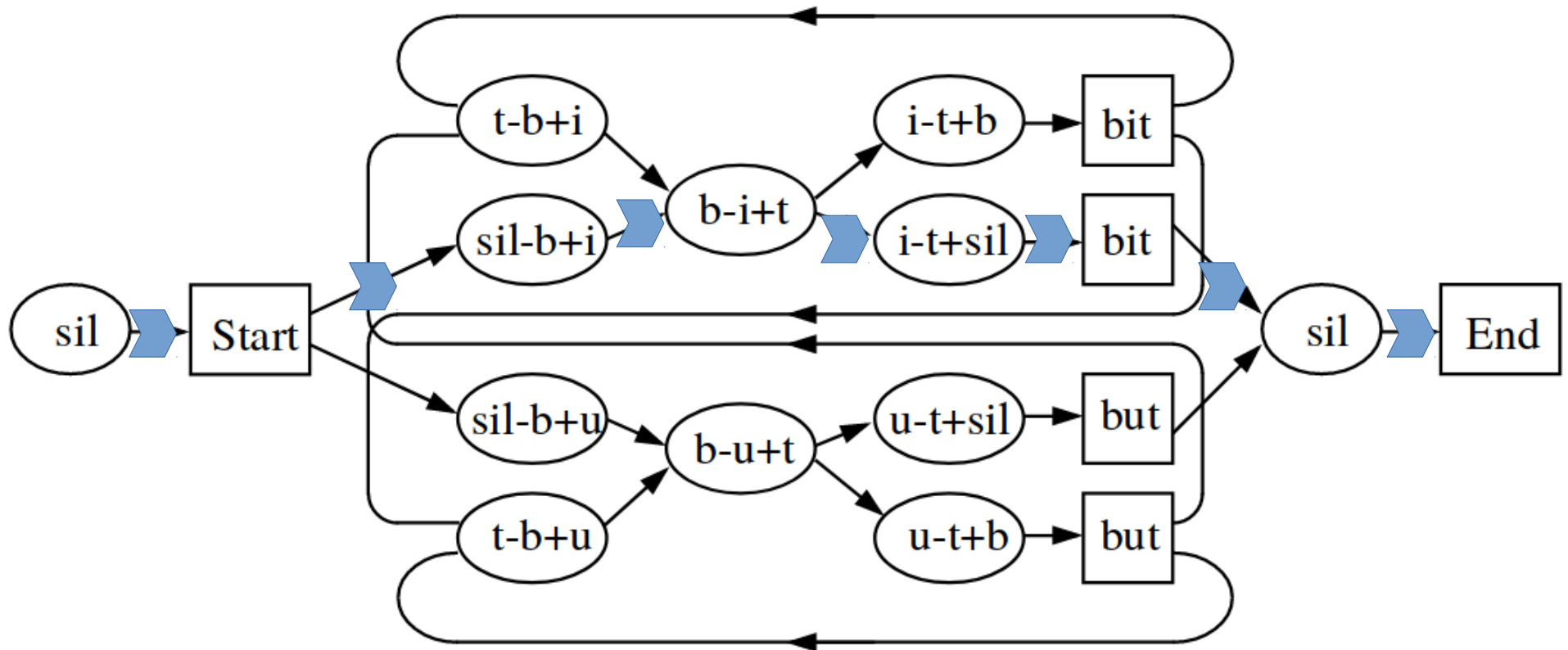


Figure credit
The HTK Book

FEATURE EXTRACTION

- Sliding window feature extraction

GMM MONOPHONE TRAINING

- Flat-start
- Baum-Welch re-estimation

GMM TRIPHONE TRAINING

- Phonetic decision tree
- Baum-Welch re-estimation

DNN TRAINING

- Labeled Data
- Gradient descent via backprop

DECODING

- WFST Graph compilation
- Viterbi decoding

Thanks for your attention:)

@joshmeyerphd

jrmeyer.github.io