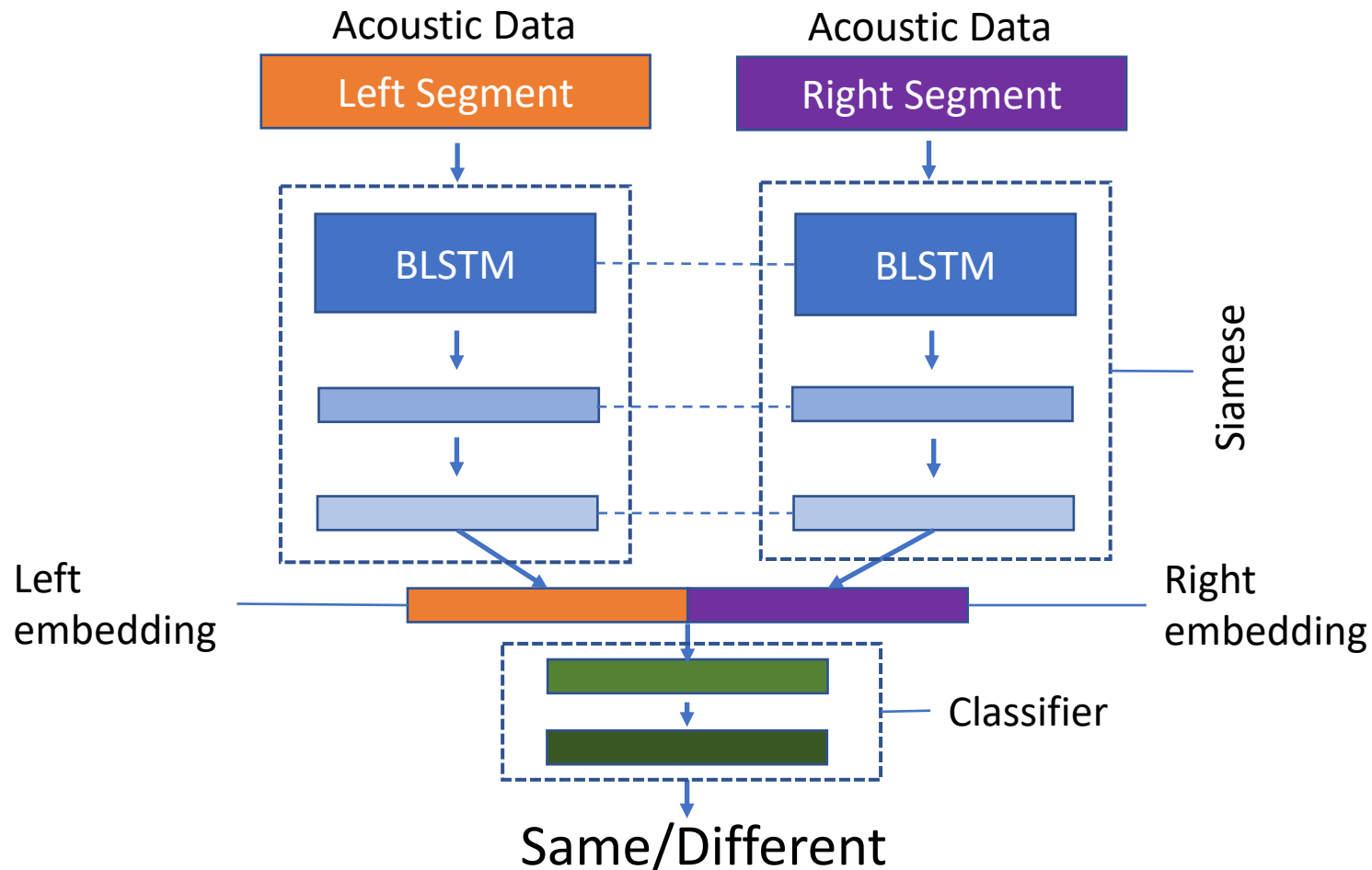


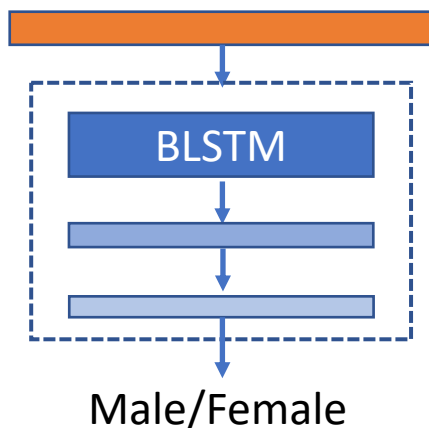
# Speaker Change Detection using Siamese Networks



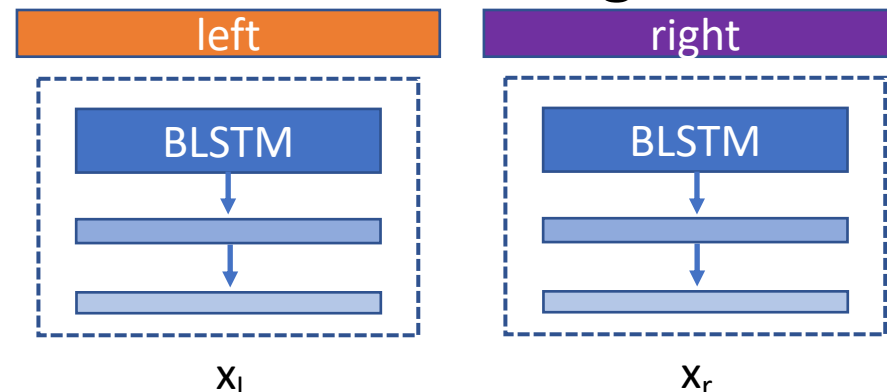
- Siamese layers share their weights
- Classifier is trained using binary cross-entropy
- Input features are PLPs

# Pre-training of the Siamese Layers

- Gender classification

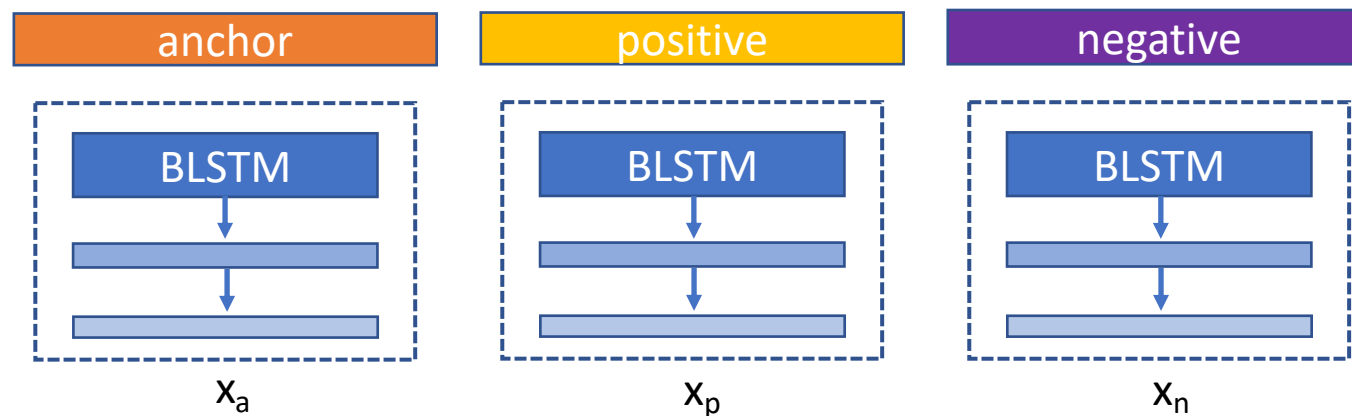


- Contrastive Divergence



$$\min \sum_{i=1}^I \delta[l = r] d(x_l^{(i)}, x_r^{(i)}) + \delta[l \neq r] \max(0, \Delta - d(x_l^{(i)}, x_r^{(i)}))$$

- Triplet Loss



$$\min \sum_{i=1}^I \max (0, \Delta + d(x_a^{(i)}, x_p^{(i)}) - d(x_a^{(i)}, x_n^{(i)}) )$$

# Validation Data Classification Accuracy (%)

Pretraining	Distance	Freeze Siamese layers	Accuracy
Gender classification	-	Yes	76.9
Gender classification	-	No	78.1
Contrastive divergence	Cosine	Yes	76.7
Contrastive divergence	Cosine	No	87.3
Contrastive divergence	Euclidean	Yes	77.4
Contrastive divergence	Euclidean	No	87.5
Triplet loss	Cosine	Yes	84.6
Triplet loss	Cosine	No	87.9
Triplet loss	Euclidean	Yes	82.7
Triplet loss	Euclidean	No	89.0