```
┌─────────────────────────────────────────────────────┐
│ Input: spectrogram N×400 (T sec × 4000Hz)           │
└─────────────────────────────────────────────────────┘
                          ↓
┌─────────────────────────────────────────────────────┐
│ Frequency Convolution: 6 filters of 1×12            │
└─────────────────────────────────────────────────────┘
                          ↓
┌─────────────────────────────────────────────────────┐
│ Frequency Max-Pooling: 1×2                          │
└─────────────────────────────────────────────────────┘
                          ↓
┌─────────────────────────────────────────────────────┐
│ Frequency Convolution: 10 filters of 1×8            │
└─────────────────────────────────────────────────────┘
                          ↓
┌─────────────────────────────────────────────────────┐
│ Frequency Max-Pooling:  1×2                         │
└─────────────────────────────────────────────────────┘
                          ↓
┌─────────────────────────────────────────────────────┐
│ Timing GRU: bi-directional, 128×2                   │
└─────────────────────────────────────────────────────┘
                          ↓
┌─────────────────────────────────────────────────────┐
│ Dense Layer: length 64                              │
└─────────────────────────────────────────────────────┘
                          ↓
┌─────────────────────────────────────────────────────┐
│ Dropout: length 64                                  │
└─────────────────────────────────────────────────────┘
                          ↓
┌─────────────────────────────────────────────────────┐
│ Soft-max: length 4                                  │
└─────────────────────────────────────────────────────┘
                          ↓
┌─────────────────────────────────────────────────────┐
│ Output: 4 posterior probabilities                   │
└─────────────────────────────────────────────────────┘
```