

# A SMALL MODEL FOR BIG ARTICULATORS

## AUTOMATIC AMERICAN SIGN LANGUAGE DETECTION WITH A TINY ML MODEL

Frederick Chan, Gina-Anne Levow, & Qi Cheng  
University of Washington Linguistics

JUST  
**1,013**  
PARAMS!



Scan & read  
or contact  
fredchan@uw.edu

### CLIP 1000s OF VIDEOS FOR YOUR DICTIONARY OR EXPERIMENT

#### UNEDITED VIDEO

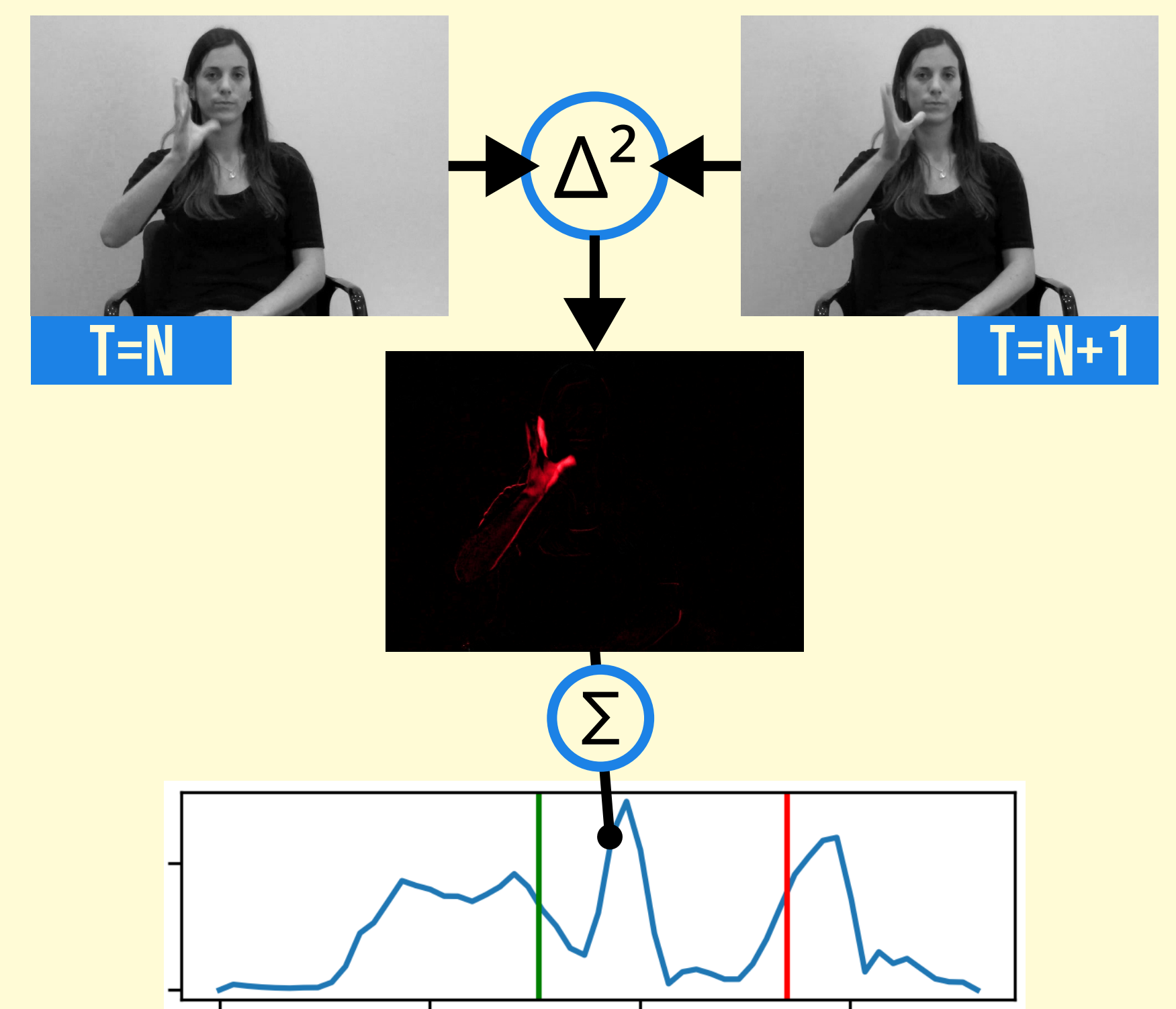
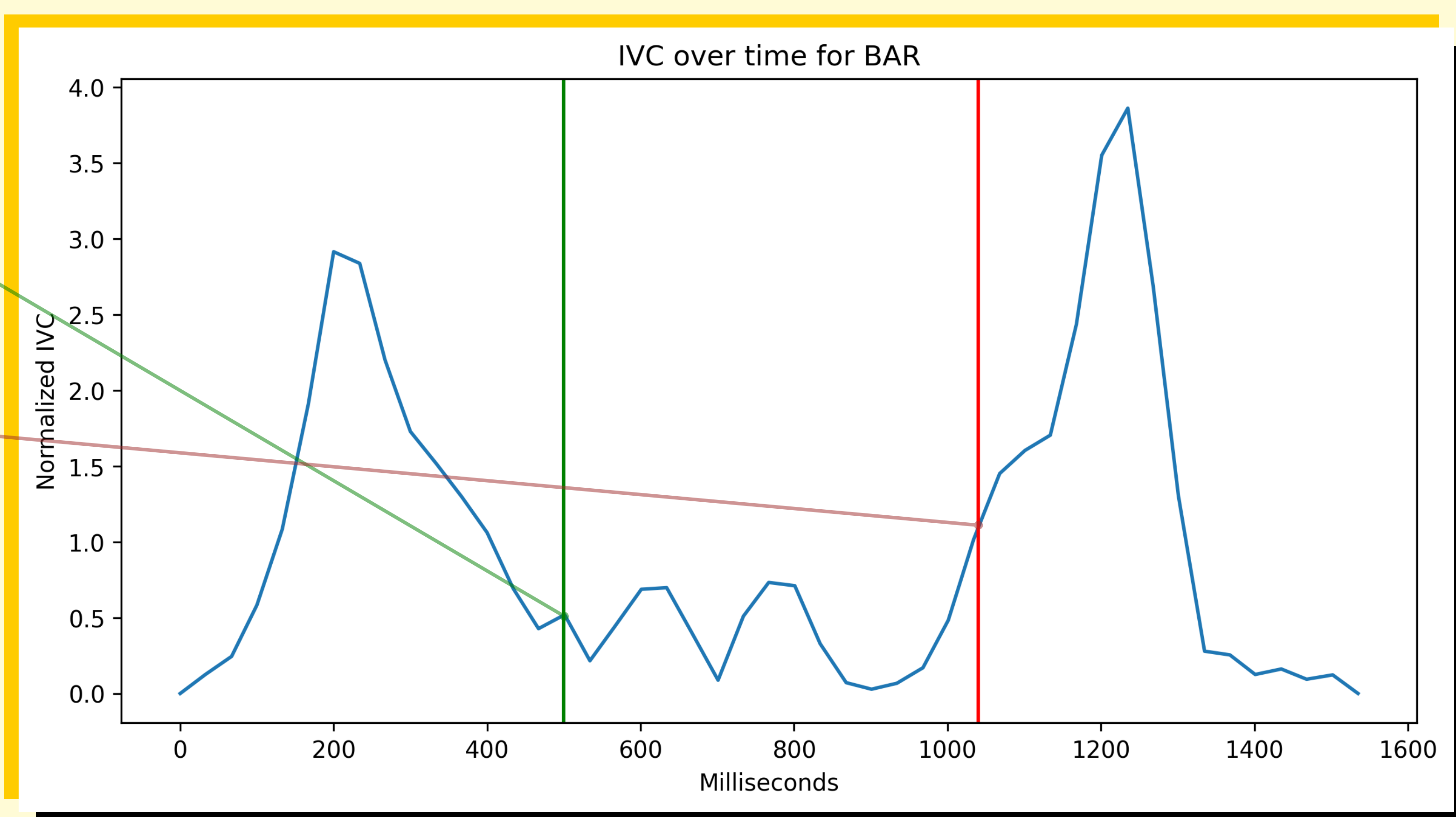
Signer records a long string of individual signs with gaps in between

#### PREPROCESSING

Automatically downsampled to 29.97 FPS, IVC computed, then normalized

#### MODEL DETECTS SIGN FRAMES

Middle frame in window of 16 frames categorized as signing/non-signing



#### INSTANTANEOUS VISUAL CHANGE (IVC)

Comparable to volume in spoken language audio, Brookshire et al. (2017) links the entrainment of brain waves to IVC in sign videos, and hypothesize that babies detect the quasiperiodic nature of IVC to "tune in" to the linguistic signal.

### TRAINED TO BE PRECISE FOR PSYCHOLINGUISTICS

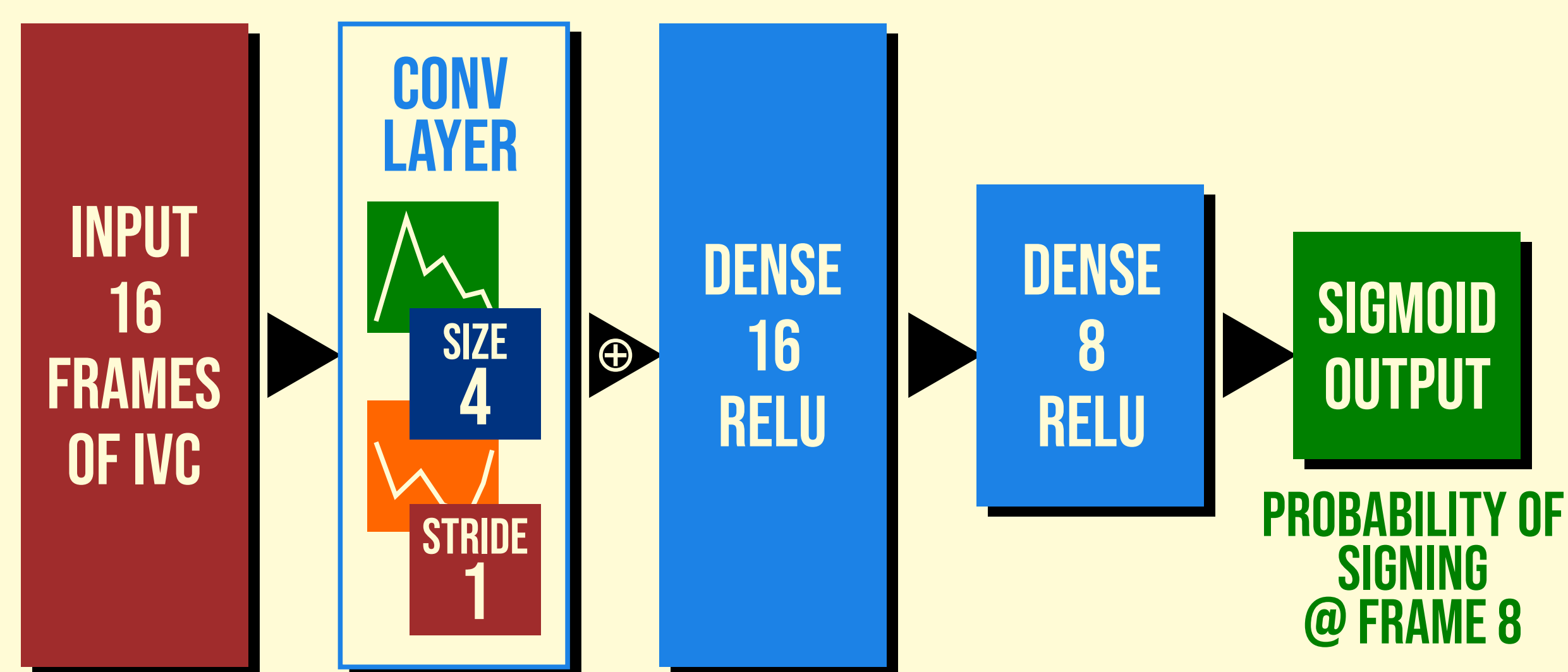
#### DEFINITIONS FROM ASL-LEX 2.0

##### ONSET

First frame where the fully-formed handshape has contacted the body or arrived at the target location near the body/neutral space

##### OFFSET

Last frame where the hand has contacted the body or before the hand(s) have begun to transition to a resting position



#### F1 SCORE

**86%**

REAL SIGNS 88%  
PSEUDOSIGNS 86%  
CONCATENATED 84%

vs IVC + Logistic Regression: 70%  
vs SoTA: 86% (indirect comparison)

#### IOU

**78%**  $\sigma=13\%$

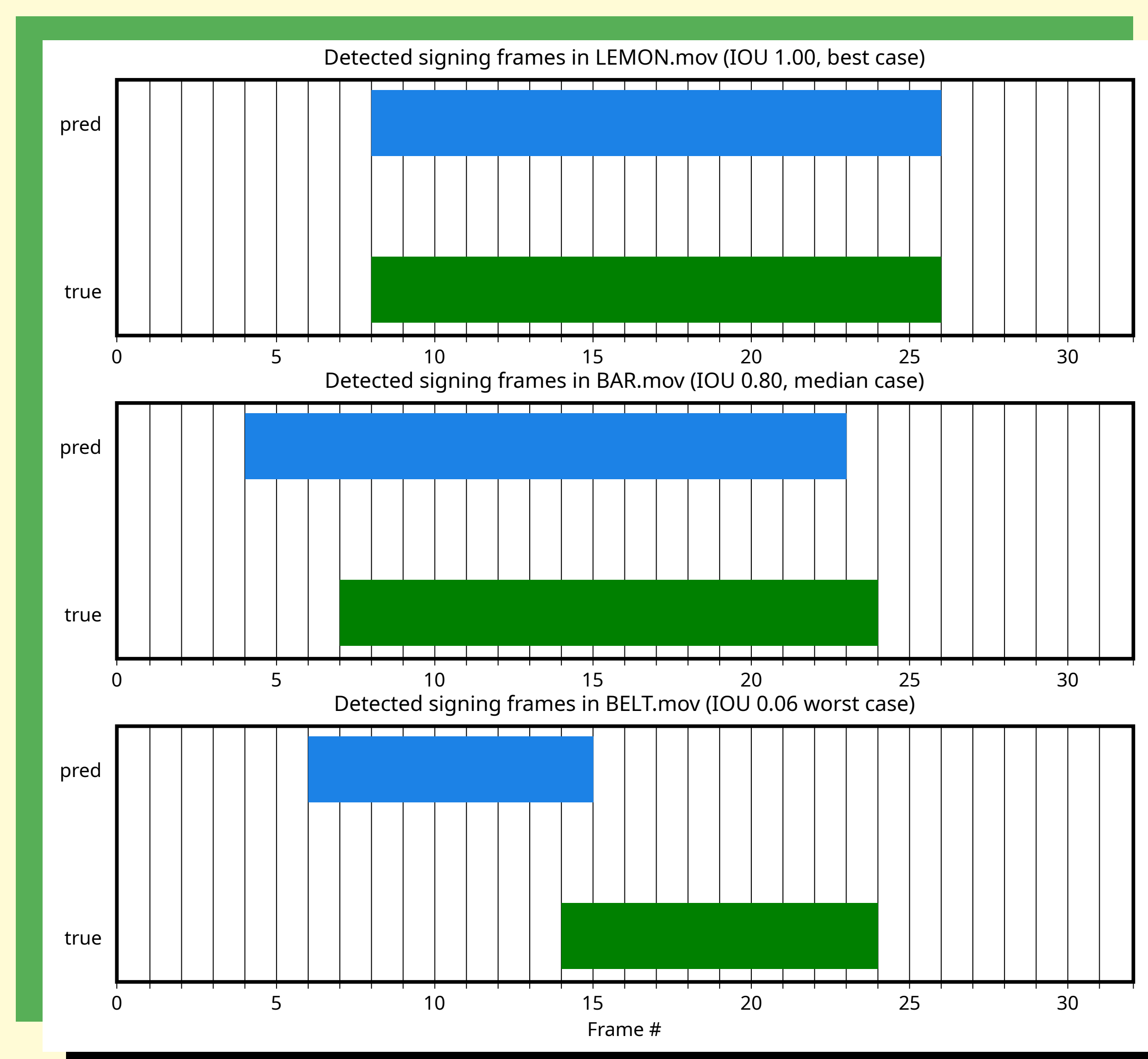
REAL SIGNS 79%  
PSEUDOSIGNS 77%  
CONCATENATED 72%

vs IVC + Logistic Regression: 64%  
vs SoTA: 77% (indirect comparison)

#### MEAN ONSET ERROR

**2.82 FRAMES**

(~90 MS)  $\sigma=2.97$  FRAMES



#### TRAIN/VALID: ASL-LEX 2.0

Sign corpus videos  
2,792 Videos  
0.86 to 6.3 sec/video  
90% Onset agreement  
99% Offset agreement



#### TEST: CASELLI ET AL 2021

Psycholinguistics stimuli  
569 Videos  
1.2 to 2.6 sec/video  
Both real & pseudosigns!

