



11th IEEE International Conference on
Automatic Face and Gesture Recognition

FG2015

My 10 Year Research Vision: Automatic Sign Language Recognition

Lale Akarun

Cihan Camgöz, Alp Kindiroğlu, Oya Aran, Cem Keskin

Computer Engineering Department
Murat Saraçlar, Electrical Engineering

Sumru Özsoy, Linguistics

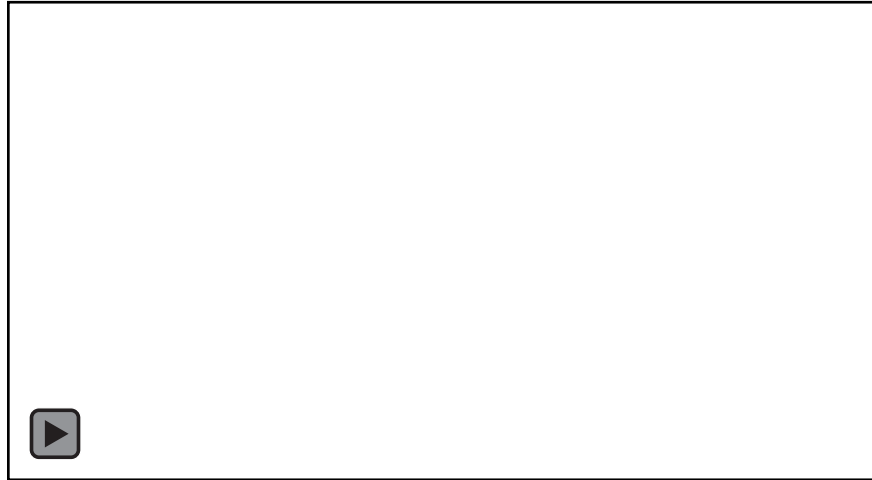
Bogaziçi University, Istanbul, Turkey



What is Sign Language?

- Fingerspelling: written language spelled in fingersigns
- Signed Speech: Words signed individually in the order of spoken language
- Sign Language: A different language from the spoken language: (ASL, BSL)

Finger Spelling

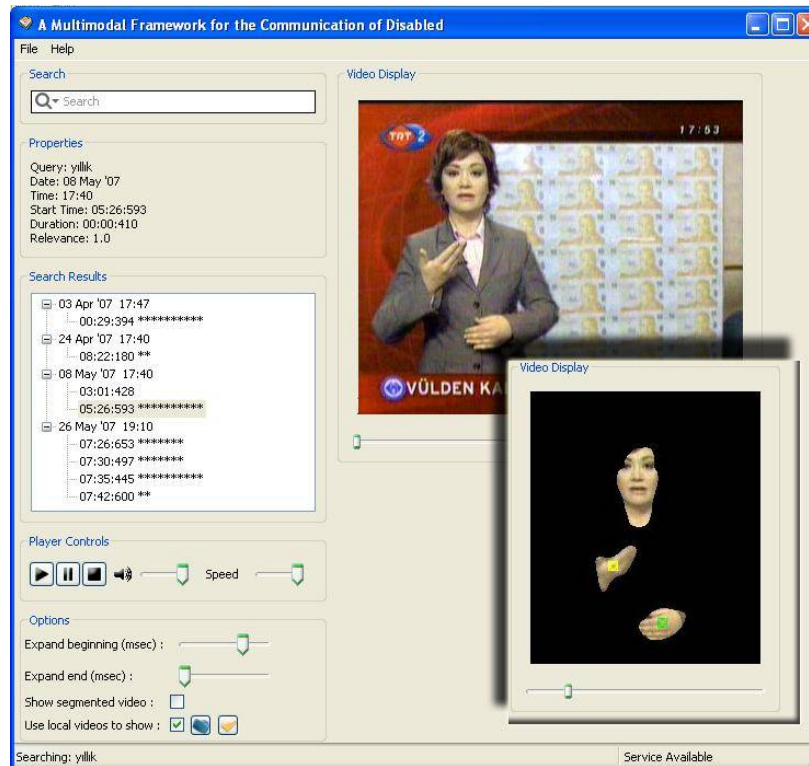


written language spelled in fingersigns

What is Sign Language?

- Fingerspelling: written language spelled in fingersigns
- Signed Speech: Words signed individually in the order of spoken language
- Sign Language: A different language from the spoken language: (ASL, BSL)

Signed Speech



News for hearing impaired: Speech, signs and sliding text

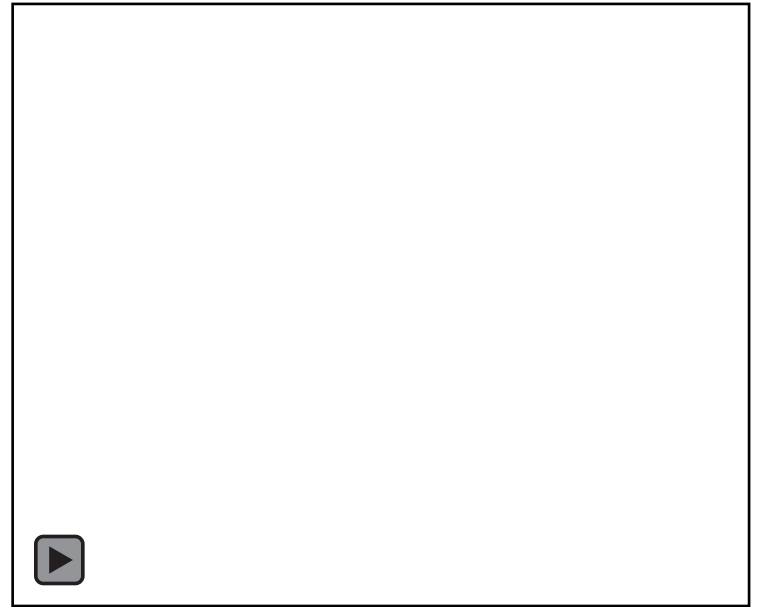
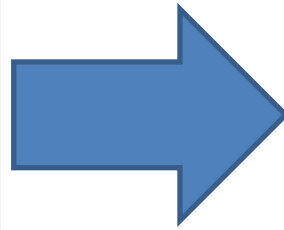
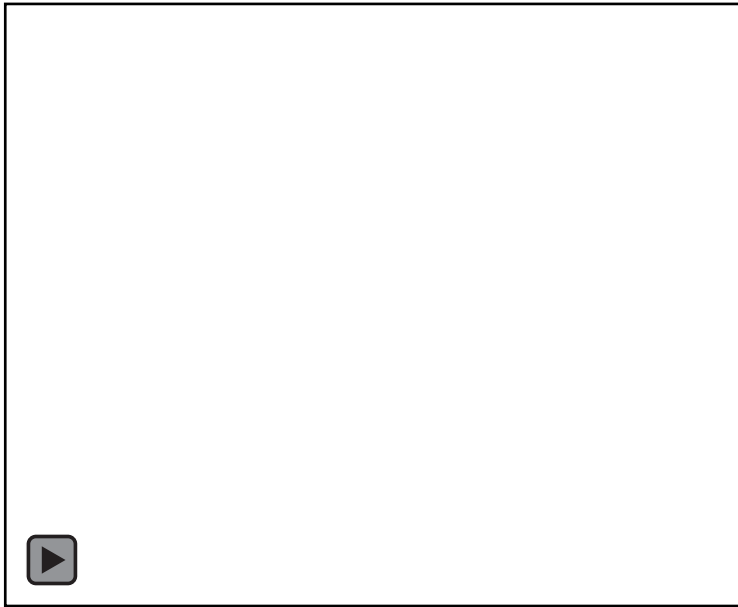
What is Sign Language?

- Fingerspelling: written language spelled in fingersigns
- Signed Speech: Words signed individually in the order of spoken language
- Sign Language: A different language from the spoken language: (ASL, BSL)

Sign Language



Question – Answer Examples



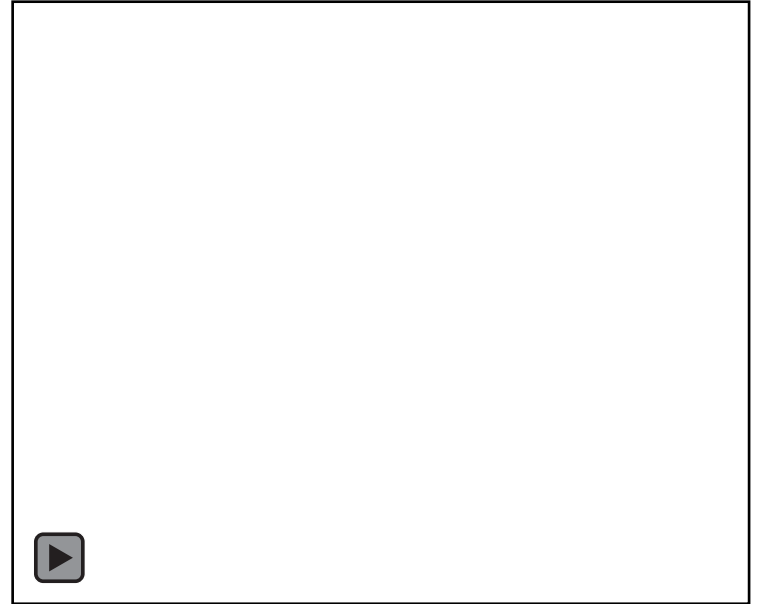
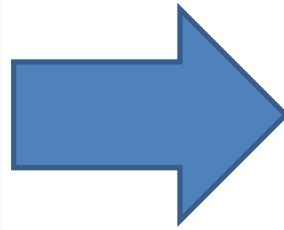
Those shoes are new, are they not?

No, these shoes are not new.

Question – Answer Examples



Your dress is lovely, enjoy it!



Thank you!

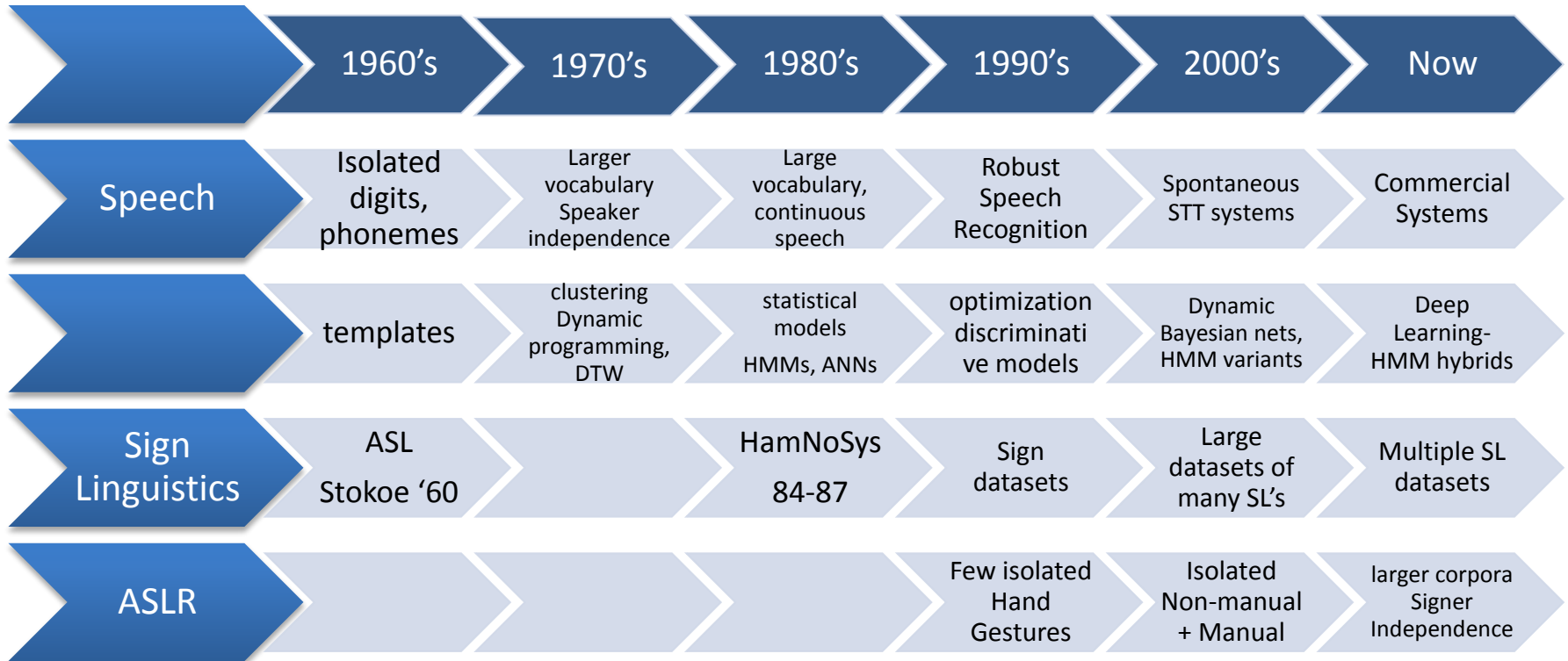
Why is Sign Language Interesting?

- Intellectual Challenge: Most basic form of human communication – sign before speech
- Combines many modalities:
 - Hand shape, hand gesture
 - Body gesture
 - Facial expressions
 - Gaze
 - Lip shape
- It is difficult: Many sign languages; cultural context
- Yet, it is intuitive for humans

Parallels with Speech Recognition

- A proper language
 - “phonemes”
 - “Co-articulation”
 - Limited vocabulary – isolated – speaker independent
-
- More difficult and in its infancy

Brief History of ASR vs. ASLR



Starner, Pentland 95
Vogler Metaxas 97

Currently Available Sign Corpora

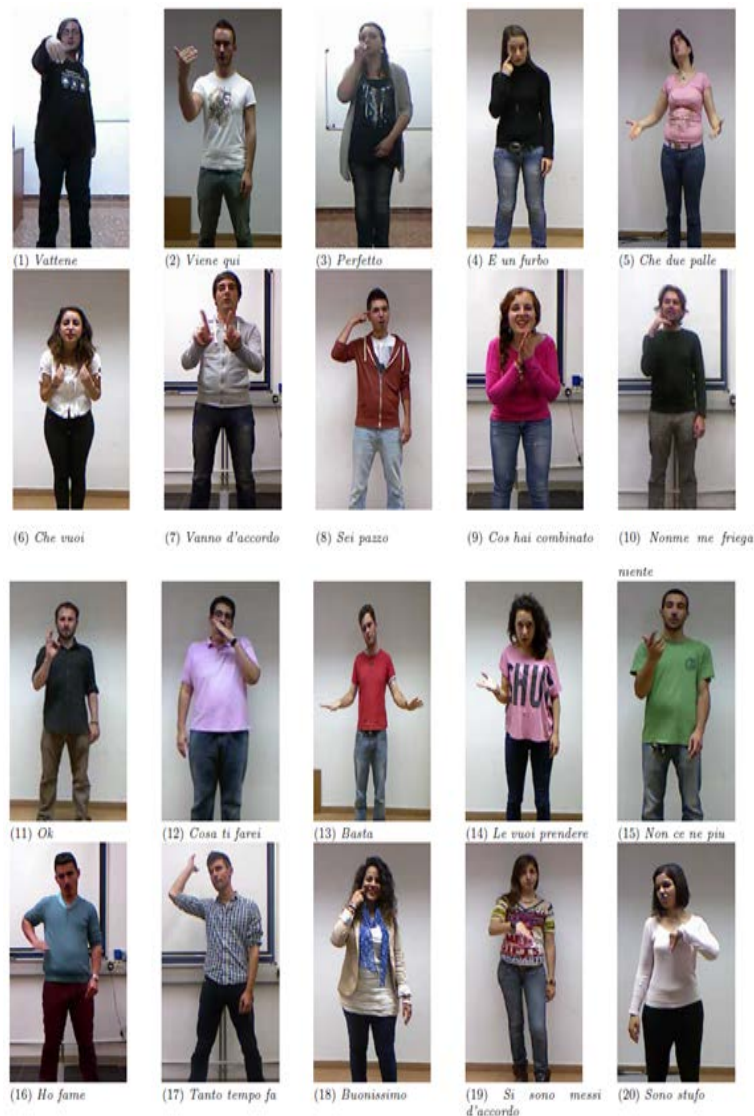
Sign Language	Project Name (Reference)	Number of		Raw Data	Annotations		
		Signs	Signers		Transcription	Manual	Non-manual
German Sign Language	DGS Corpus (2009-2023) Hanke et. al. 2009	~6000 Unique Signs (~3,000,000 in total)	330	HD Videos: 3x720p & 2x1080p	Glosses & HamNoSys	Token, Deviation, Citation Form	Mouthing, Facial Expression
Sign Language of the Netherlands	Corpus NGT (2006-2008) Crasborn et. al. 2008	2375 Video Clips	92	HDV and DV	Glosses	Translation	Head Shakes, Mouth Actions
Australian Sign Language (Auslan)	Endangered Language Documentation Project Johnston et. al. 2010	4518 Video Clips	100	Digital Videos	Glosses (IDglosses)	Auslan Annotations	
British Sign Language (BSL)	British Sign Language Corpus Project Schembri et. al. 2014	Over 2,500 Signs (50,000 Sign Tokens)	249	Digital Videos	Glosses (IDglosses)	Lexical Meaning, Hand Shape and Orientation	N/A
Multiple Sign Languages (British, German, Greek, French)	Dicta-Sign Matthes et.al. 2012	16-18 hours of videos per signer	14-18 signers per language	7 Cameras: 2 of them Stereo Cameras	HamNoSys	Sign Level, Content and Clause Boundaries Tagging.	Mouthing
						Some of the Auslan Annotations.	
American Sign Language	ASLLRP - ASLLVD Neidle et. al. 2012	>3,300 ASL Sign 9,800 tokens	6	4 Cameras with different angle and resolutions	Glosses	Sign Start and Stop Times, Start and Stop Hand Shape, Morphological Annotations, Articulatory classifications	N/A
American Sign Language	Conly et. al. 2013	1313 (3000 when complete)	2	Kinect Color and Depth 25fps 640x480	Glosses	Hand Position, One/Two Handed	N/A

Sensors

- Instrumented gloves
- Accelerometers
- Color Gloves
- Stereo Camera
- Depth Camera
- Kinect

ChaLearn 2014 – Gesture Challenge Dataset

- Continuous Gesture Spotting
- ChaLearn 2014 LAP Track 3 Dataset
 - 20 Italian Cultural/Anthropological Gestures
 - More than 14,000 Gestures
 - Abundance of Data (Suitable for complex learning methods)



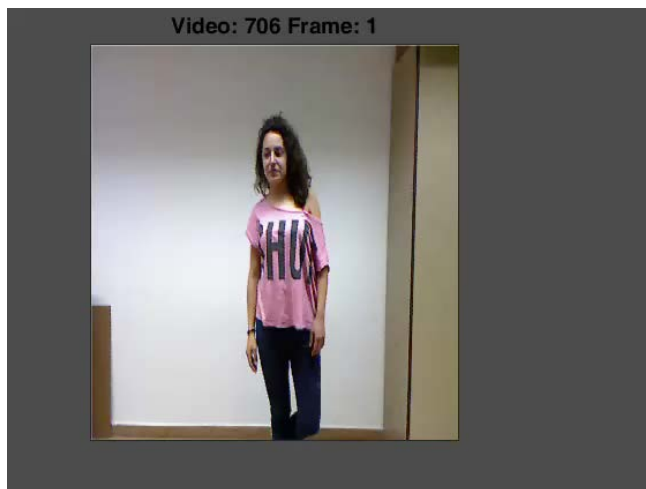
ChaLearn 2014 – Our Sample Result



Good Recognition Example



Gesture Stance in Non-Gesture Segments



Co articulated Gestures



Out of Dictionary Gestures

Bosphorus Sign - Turkish Sign Language Corpus for Health and Finance Domains

- It will be recorded using Kinect for Windows v2. Raw Data Modalities will be:
 - Color video (1080x1920)
 - Depth video (424x512)
 - Skeleton Joints



Bosphorus Sign - Turkish Sign Language Corpus for Health and Finance Domains

- 674 unique signs from different domains
 - 57 General Words (i.e.: Family, Friend, Mother, Father, Car, etc...)
 - 328 Health Terms (i.e.: Liver, Pediatrician, Bleeding, Heart Attack, Bandage, MRI, etc...)
 - 198 Finance Terms (i.e.: Check, Deposit, Withdraw, Limit, Bill, Account, etc...)
 - 91 Verbs (i.e.: Eat, Sleep, Listen, Walk, Run, Jump, etc...)



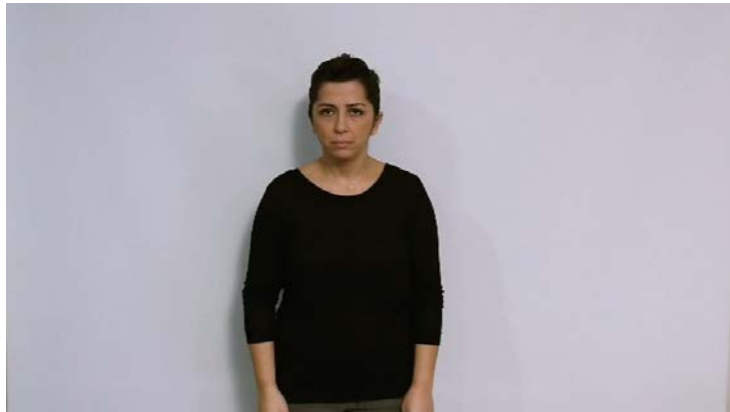
Pain reliever



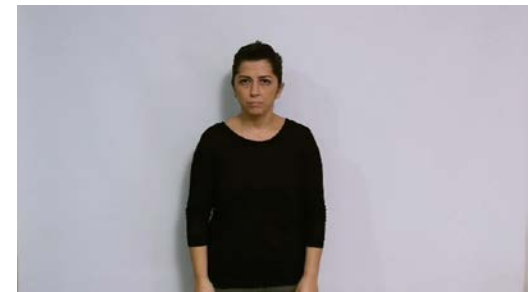
Emergency room”

Bosphorus Sign - Turkish Sign Language Corpus for Health and Finance Domains

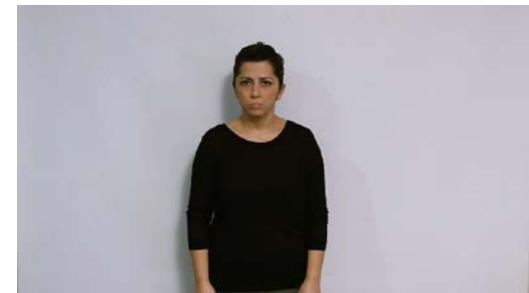
- 60 + sentences (questions and their answers) for kiosk hospital and bank applications
 - How can I help you? (I feel sick, I am here to visit a patient, ...)
 - Which kind of account would you like to open? (A Foreign Currency Account, A Private Retirement Account)
 - Do you have an appointment? (Yes I do have an appointment, No I don't have an appointment)
 - Would you like to withdraw or deposit money? (I would like to withdraw money, I would like to deposit money)



Do you have an appointment?



Yes, I have an appointment.



No, I don't have an appointment.

Hospital Kiosk Application – In English



Bosphorus Sign - Turkish Sign Language Corpus for Health and Finance Domains

- It will be collected from native signers and early learners
- Annotations will include:
 - Segmentation of sign in the video
 - Hand Shapes and orientations
 - Non-manual annotations such as: head movements, facial expressions, etc.
 - HamNoSys
- Current State: First Recordings have been done by professional sign language tutor.
- Additionally: User Friendly Kinectv2 Recording Tool and Annotation tool will be publicly available shortly after the conference.
 - Online sign segmentation and validation by recorder.
 - Compression option after the recording is done.
 - Easily configurable options such as save locations and gesture list.



Gesture Recorder



Save Location:

Gesture List:

Current Gesture

Gestures

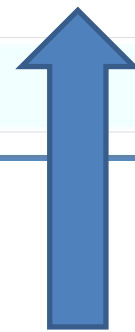
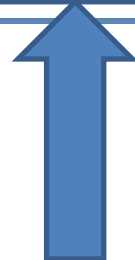
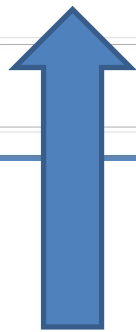
Initializing

Recording Options

- Record Depth Stream
- Record Color Stream
- Record Body Stream
- Compress Color Video (MP4)
- Compress Depth Video (MP4)

Display Options

- Color Video
- Depth Video
- Infrared Video
- User Masked Video (Moon)
- Draw Body Coordinates



Choose Your Save Location

Load Your Gesture List

Choose Your Modalities

Start Recording



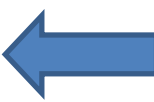
Save Location: D:\Recorder\Demo Gesture List: Health_Domain.txt Load List **How Can I Help You?** Next Gesture Start/Stop Gesture Validate Gesture Stop Recording

Gestures: How Can I Help You? Pain Emergency Emergency Service

Recording Options:
 Record Depth Stream
 Record Color Stream
 Record Body Stream
 Compress Color Video (MP4)
 Compress Depth Video (MP4)

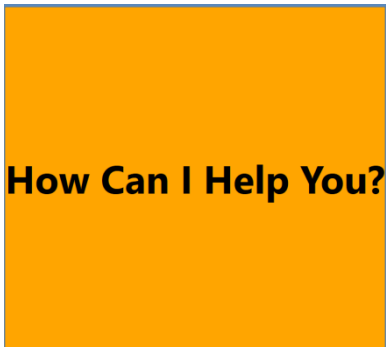
Display Options:
 Color Video
 Depth Video
 Infrared Video
 User Masked Video (Moon)
 Draw Body Coordinates

Initializing



Signer Screen
(Orange: Next Gesture)

Initiate Gesture when
the user is ready





Gesture Recorder



Save Location: D:\Recorder\Demo Gesture List: Health_Domain.txt Load List

How Can I Help You? Next Gesture Start/Stop Gesture Validate Gesture Stop Recording

Prev. Gesture

Initializing

Gestures

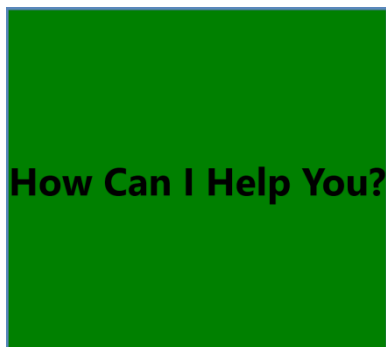
- How Can I Help You?
- Pain
- Emergency
- Emergency Service

Recording Options

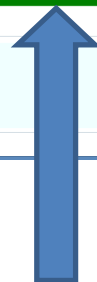
- Record Depth Stream
- Record Color Stream
- Record Body Stream
- Compress Color Video (MP4)
- Compress Depth Video (MP4)

Display Options

- Color Video
- Depth Video
- Infrared Video
- User Masked Video (Moon)
- Draw Body Coordinates



Signer Screen
(Green: Perform Gesture)



End gesture when
the user is done



Save Location: Gesture List: **How Can I Help You?**

Gestures:

Recording Options
 Record Depth Stream
 Record Color Stream
 Record Body Stream
 Compress Color Video (MP4)
 Compress Depth Video (MP4)

Display Options
 Color Video
 Depth Video
 Infrared Video
 User Masked Video (Alpha)
 Draw Body Coordinates

Initializing

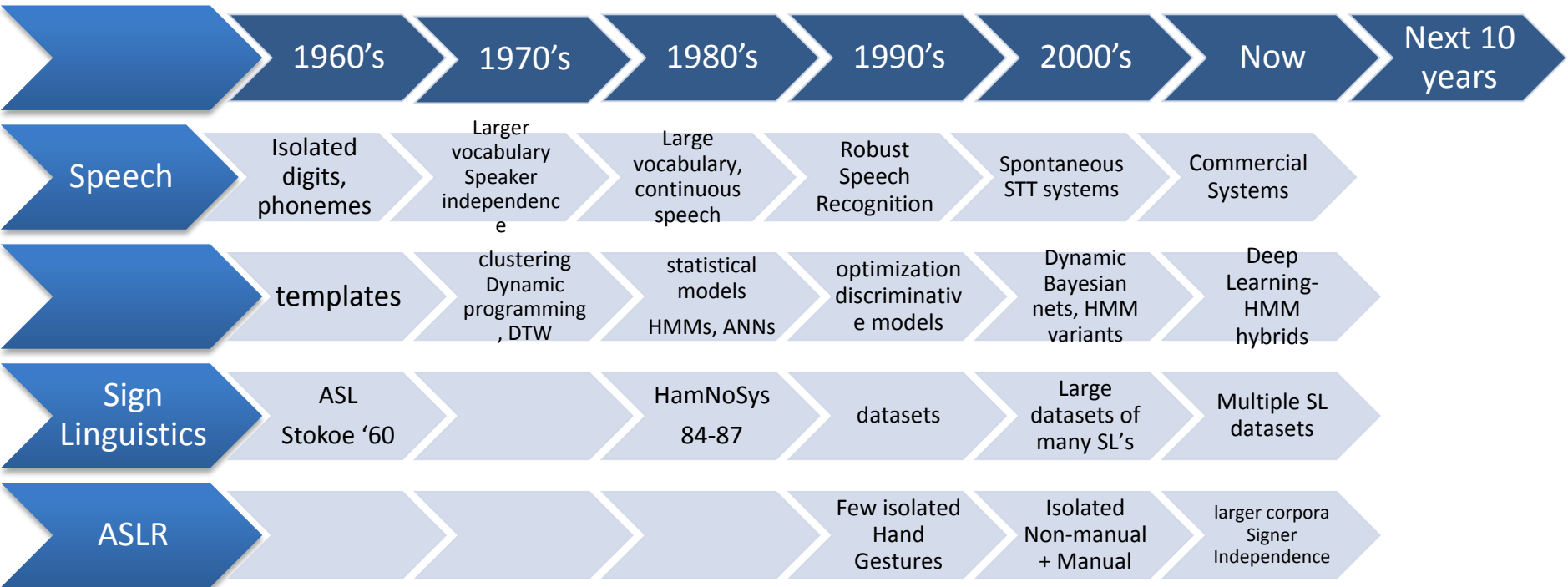


Signer Screen
(Red: Gesture is Done, Rest)

Next Gesture
Go to
Validate
Gesture



Next 10 years in ASLR



Can we follow the same path as speech?

- Multimodal; so the amount of data needed explodes.
- 4D data
- Synthetic data?

Synthetic Dataset Generation



What makes Sign more Difficult

- Different Modes have different natures
 - Hand trajectory: time series
 - Hand Shape: significant when stationary
 - Facial expression: intensity is a modifier
 - Speed of signs is a modifier for verbs
- Time Synchrony of modes very important
- 3D Space is also important
- New models, new techniques needed

What can we learn from other fields?

- Speech: Large datasets; challenges
 - More challenging challenges
- Linguistics: Better description of sign languages
- Computer animation: create labeled training data
- Social signal processing: facial expressions – can they help to segment signs?

Research Challenges: Next 10 Years

- From isolated SR to Sign spotting
- Sign segmentation
- Continuous signing
- How to make it robust?
- How to make it signer independent?
- Signers add signs all the time – how to include new signs?

Research Challenges: Next 20? Years

- Continuous Sign Language Recognition (of multiple signers)
- How to adapt to dialects?
- How to learn a new language automatically?