Face Classification with Ultrasonic Sensing

Phillip McKerrow and Kok Kai Yoong

The echo of a chirp of ultrasonic energy from an object contains information about the geometry of that object: relative depth of surfaces and approximate area of those surfaces. A human face has complex geometry that produces a distinctive echo. In this paper, we report initial research into whether there is sufficient information in the echo to recognize a face. Potential features for classification are identified using a facial model. The classification results are poor, but encourage future research to find better quality features.

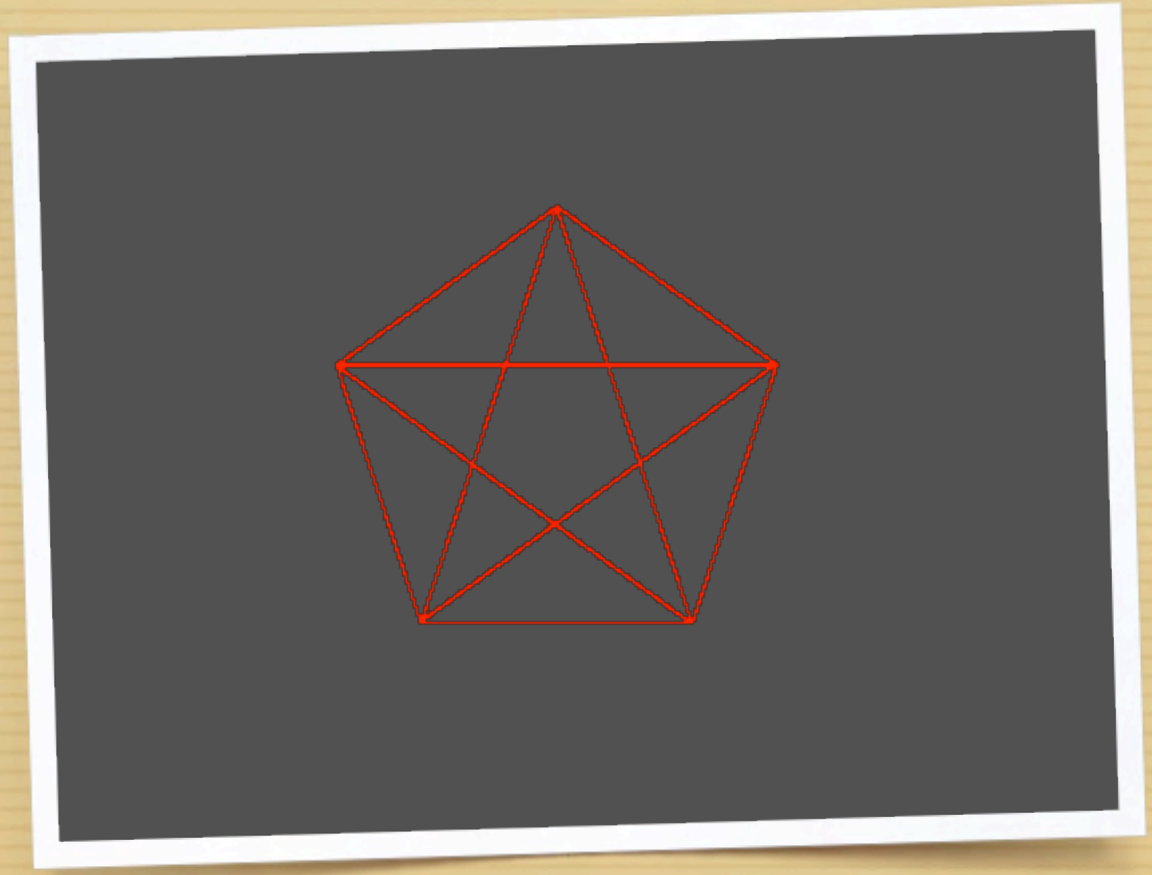
TAROS'06, Guildford, England


Wollongong

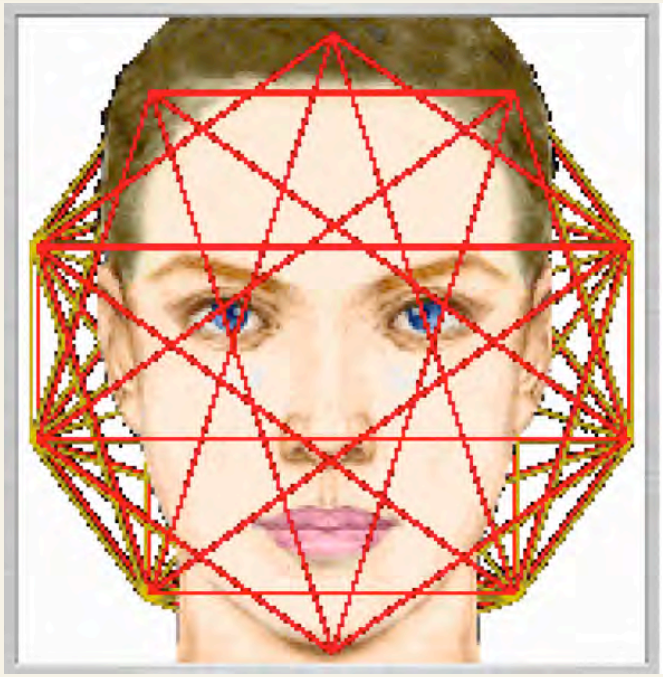
Faces

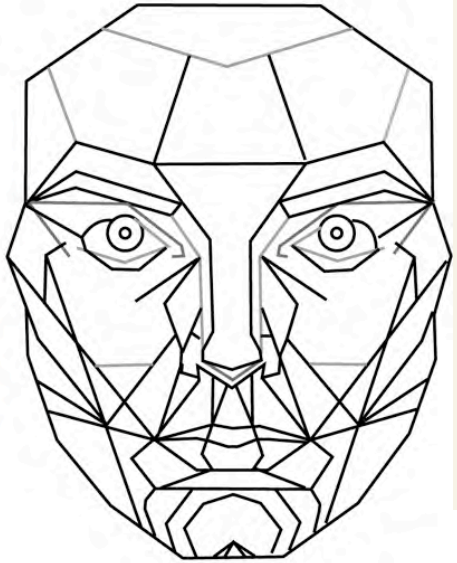
- Previous work - plants, roughness
- faces are interesting
- important in recognition
- biometric with computer vision
  - airport recognition - employees 53% correct
- is it possible to classify a face with the information in an echo of an ultrasonic chirp?
- Dror - 5 faces, 96% with Neural Network
  - didn't scale up

Marquart Beauty Analysis



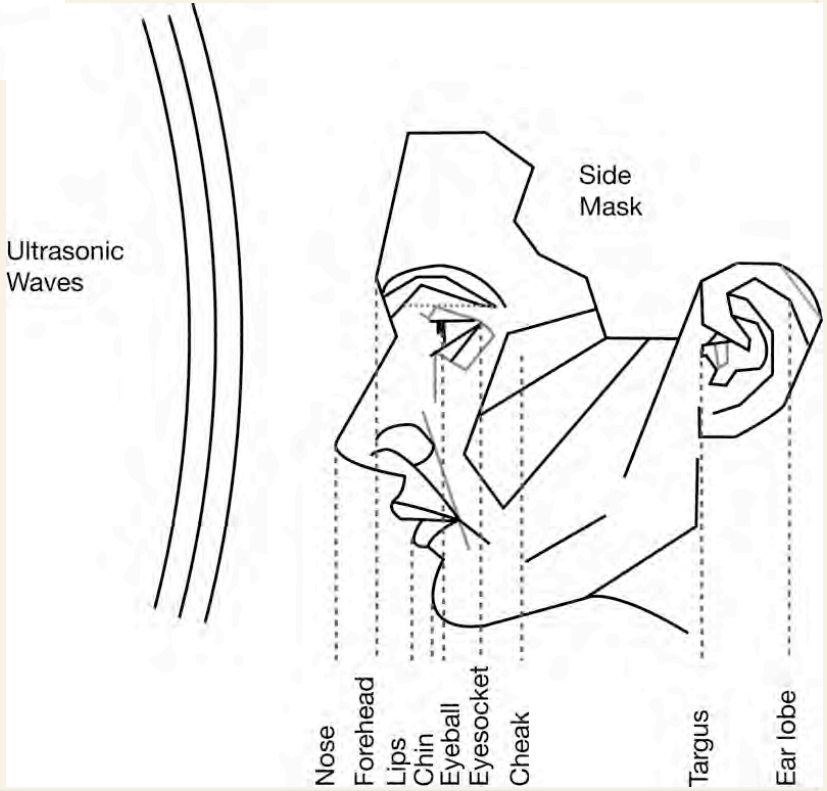
Face Geometry





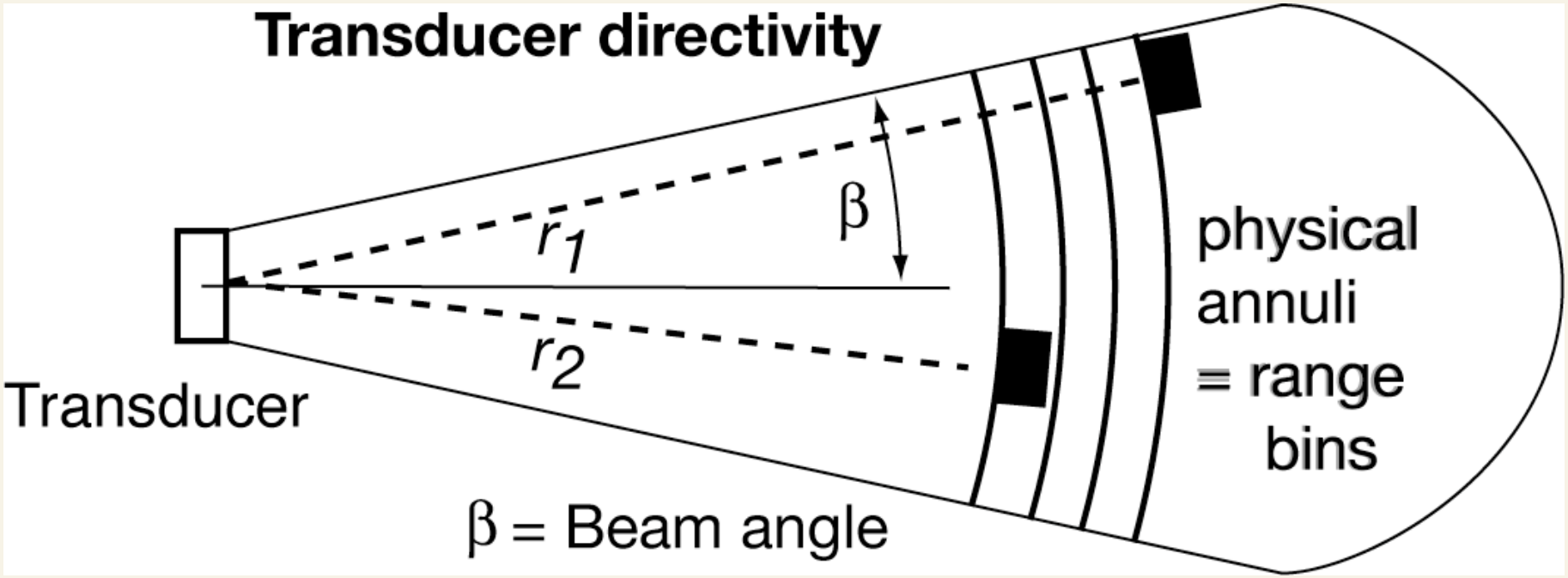
Repose Mask
Models face with flat facets
~ specular reflectors
~ 1st echo from nose

~ Ultrasonics is a range area sensor
~ Information in echo determined by geometry of object
~ time ∝ range (facet depth)
~ amplitude ∝ energy ∝ facet area

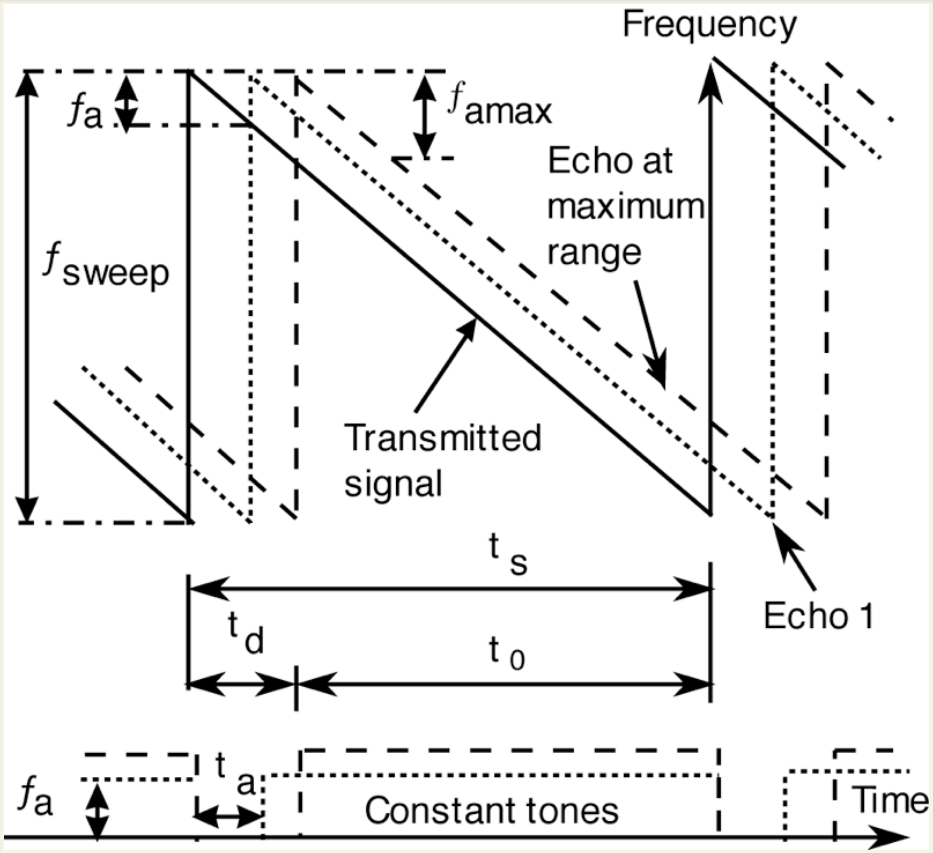


K-sonar mobility aid for blind people

CTFM Sensing

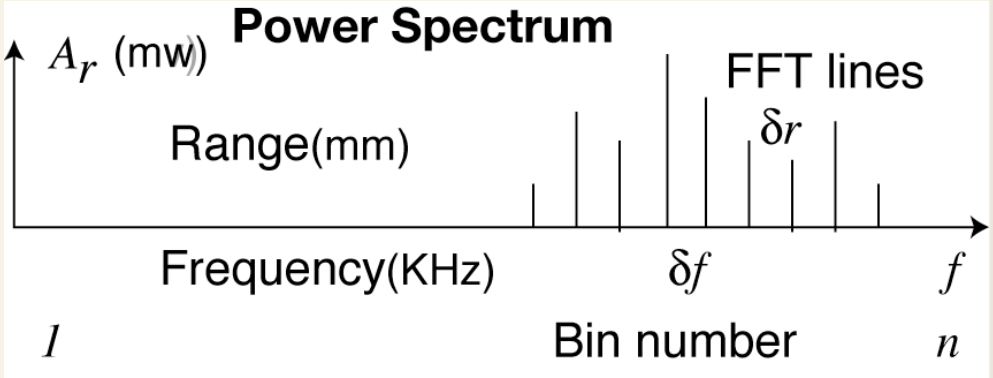


- 19mm diameter
- vertical beam angle 19.32°, horizontal 7.6°
- sweeps down from 100 to 50KHz every 100msec
- echo is delayed and filtered
- demodulate echo with transmission
- difference frequencies (0..5KHz) are ∝ to range
- fft calculates power spectrum
- converts time domain to frequency domain



- sweeps down
- echo is delayed
- demodulate echo with sweep
- range-energy signal

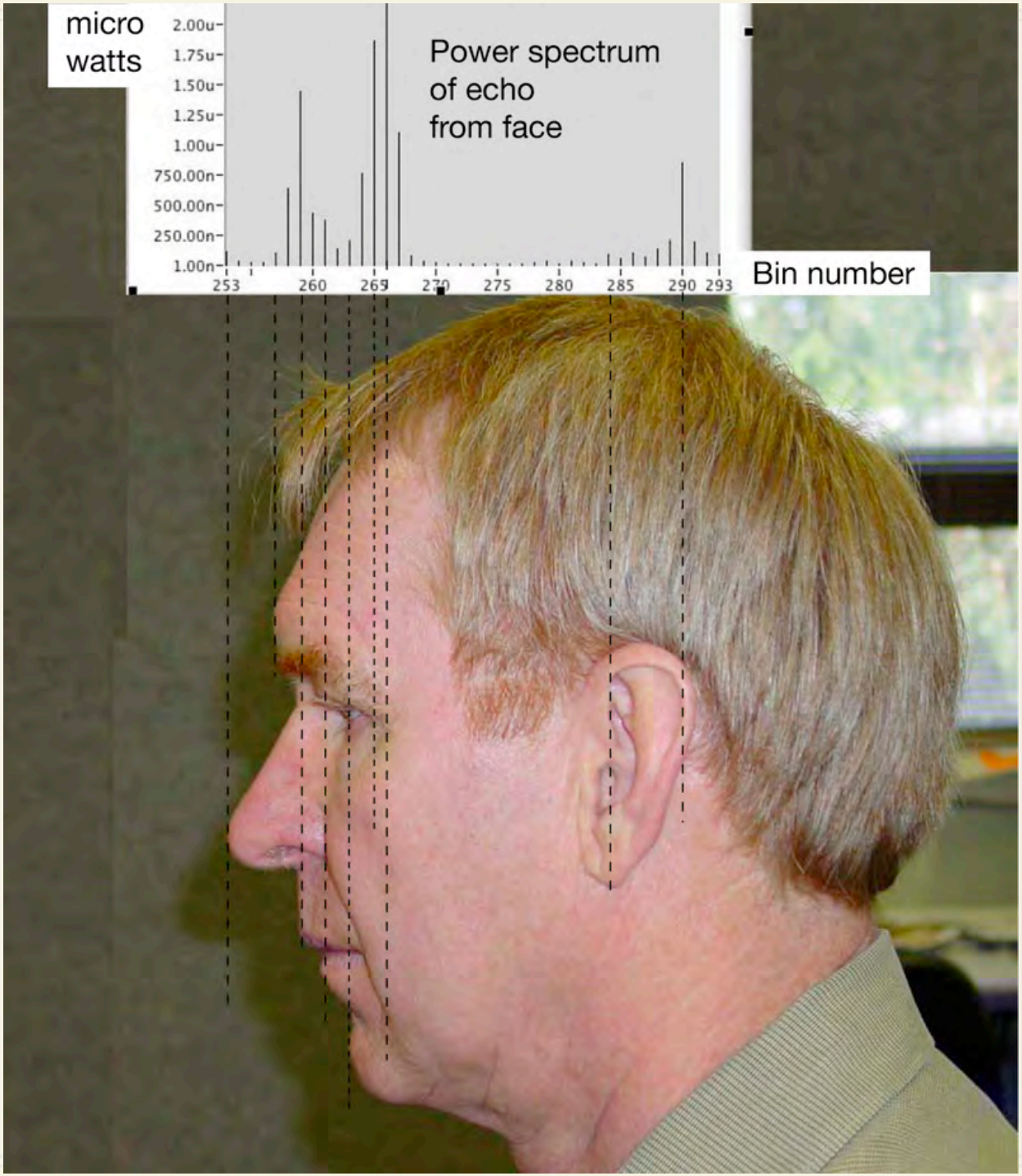
- 512 bins,
  - $\delta f = 10\text{Hz} \propto 3.87\text{mm}$
- bin i
  - frequency f_i
  - reflected from range r_i
  - energy = amplitude A_i



Relationship between echo and face depth

Side Repose Mask
~ facial features separated in depth
~ their echoes should be at different times

Power spectrum lined up with facial features



Acoustic face geometry table

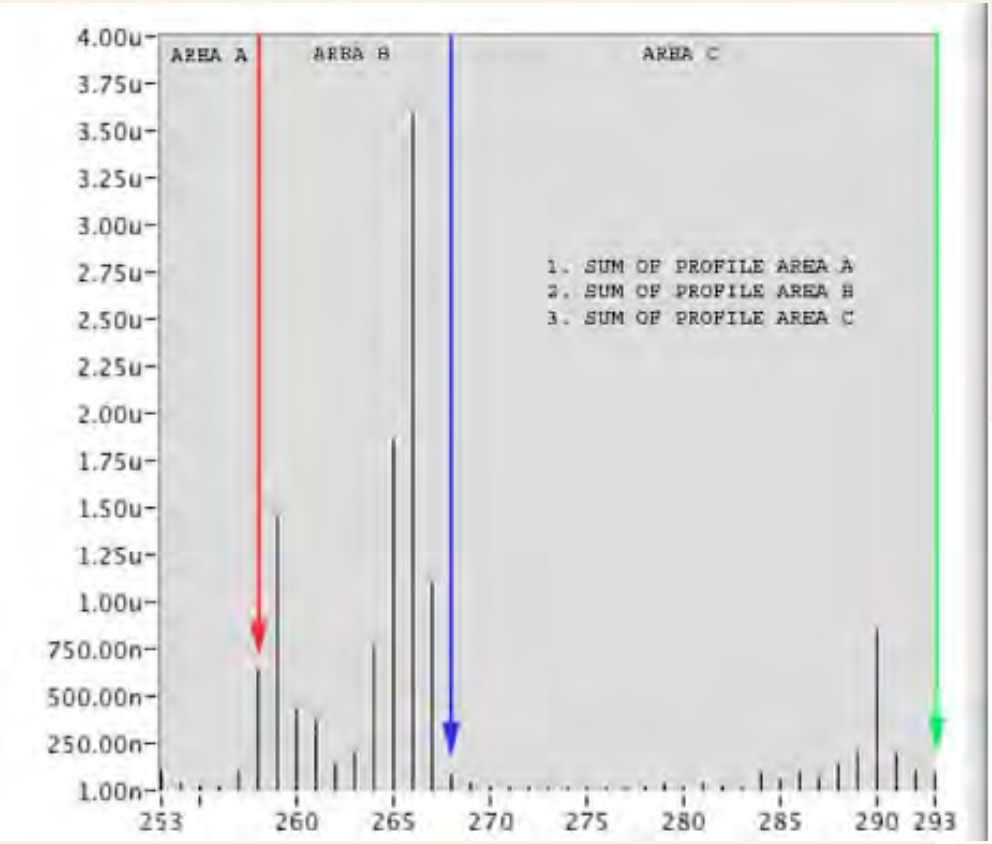
Nose	Fore head	Lips	Chn	Eye	End of Eye	Stat of Ear	Ear Lbe e	End of Ear
243	244	246	248	**248**	251	269	271	268
254	259	256	**260**	**260**	264	283	288	288
247	251	248	252	**253**	**257**	273	276	283
250	252	254	259	**256**	**259**	277	282	285
259	263	**262**	264	264	267	284	286	292
255	260	259	262	**263**	265	283	285	295
254	258	256	255	**260**	263	281	282	292
252	256	255	258	**258**	**261**	279	280	290
259	266	262	262	**266**	270	290	292	299
253	257	259	263	263	**265**	282	284	291

Facial features and echo bins for 5 males and 5 females
- red bins have highest amplitude echo component

Features

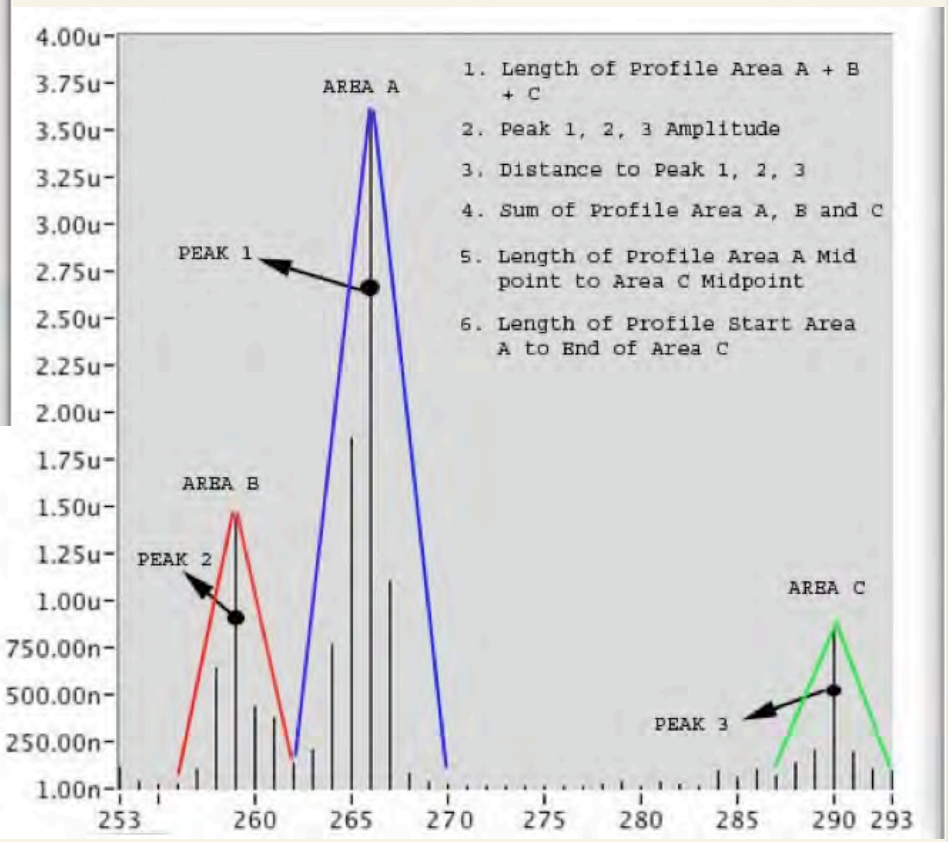
- started with features from previous research
- window the region of interest based on geometry
- length of window - acoustic density profile
  - ∝ depth of face - nose to ear lobe
- distance to peak 1 ∝ tip of nose to edge of eye
- sum of energy in profile
- face profiling into 3 regions
  - sum of energy in each region

Dividing acoustic density profile into 3 regions



Based on face model

- A. nose to chin - bin 6
- B. chin to mid face - bin 15
- C. mid face to end ear



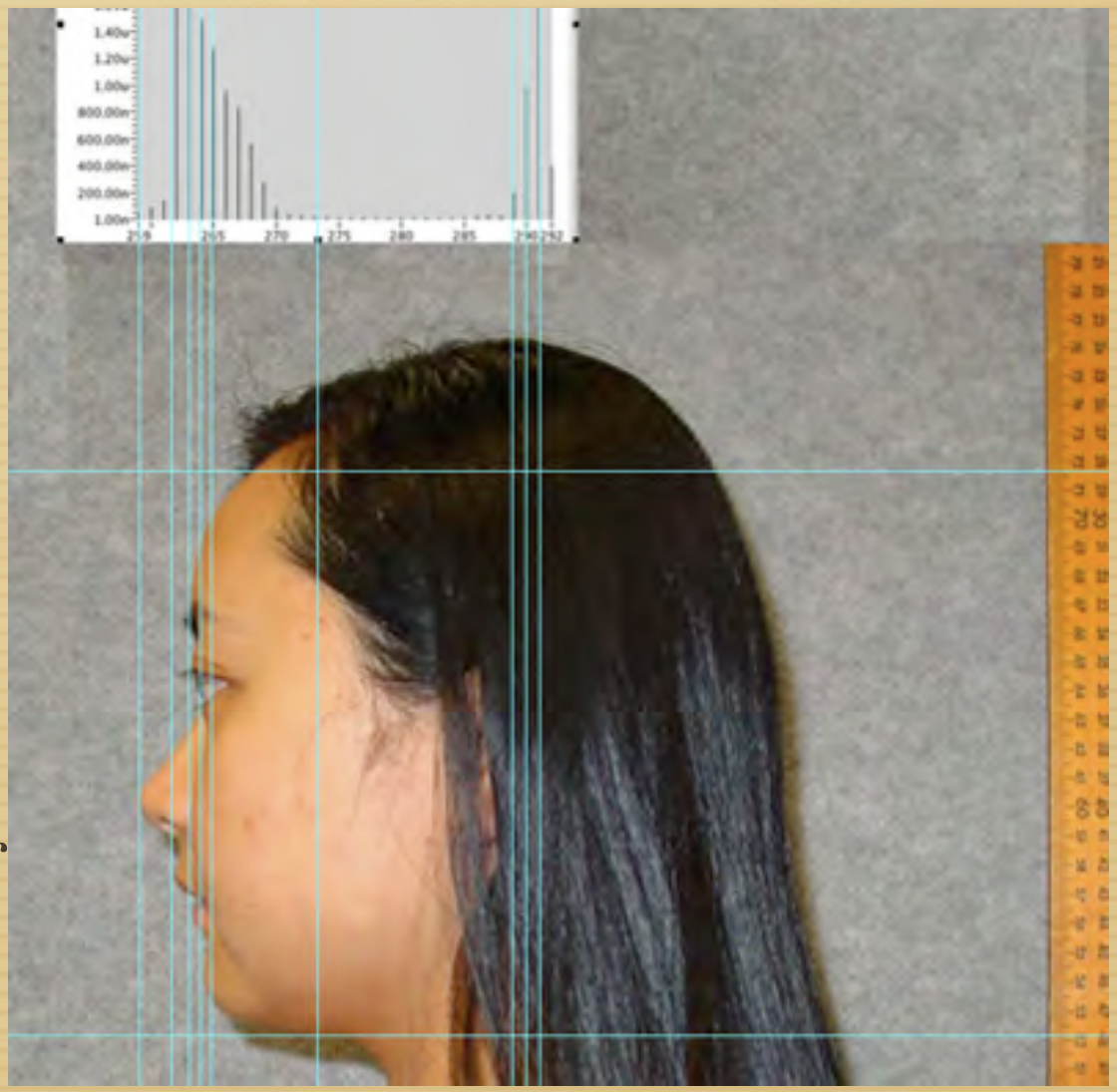
Based on signal features

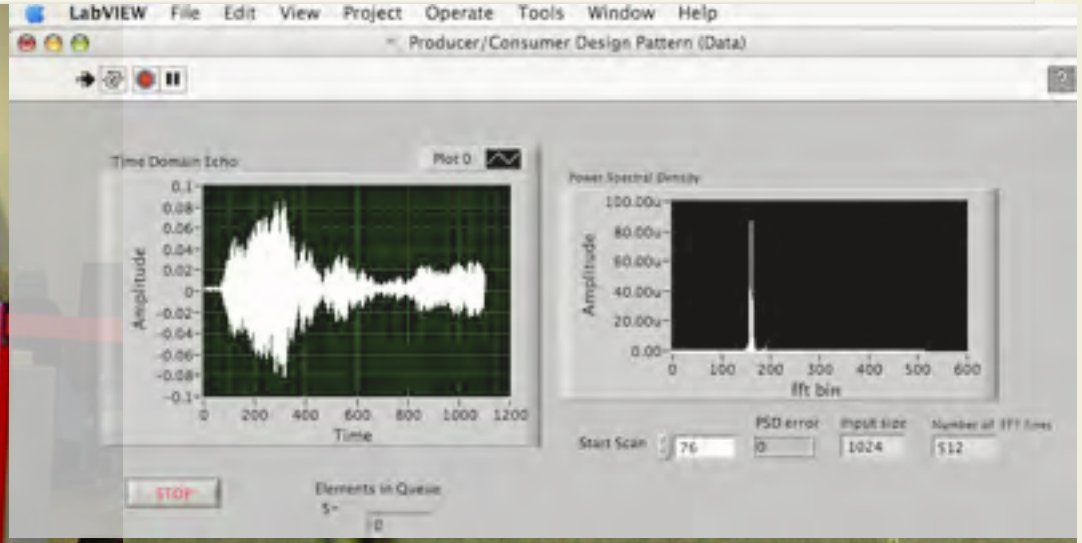
- regions around 3 peaks
  - peak 1 - flat of face
  - peak 2 - edge of eye
  - peak 3 - front of ear lobe

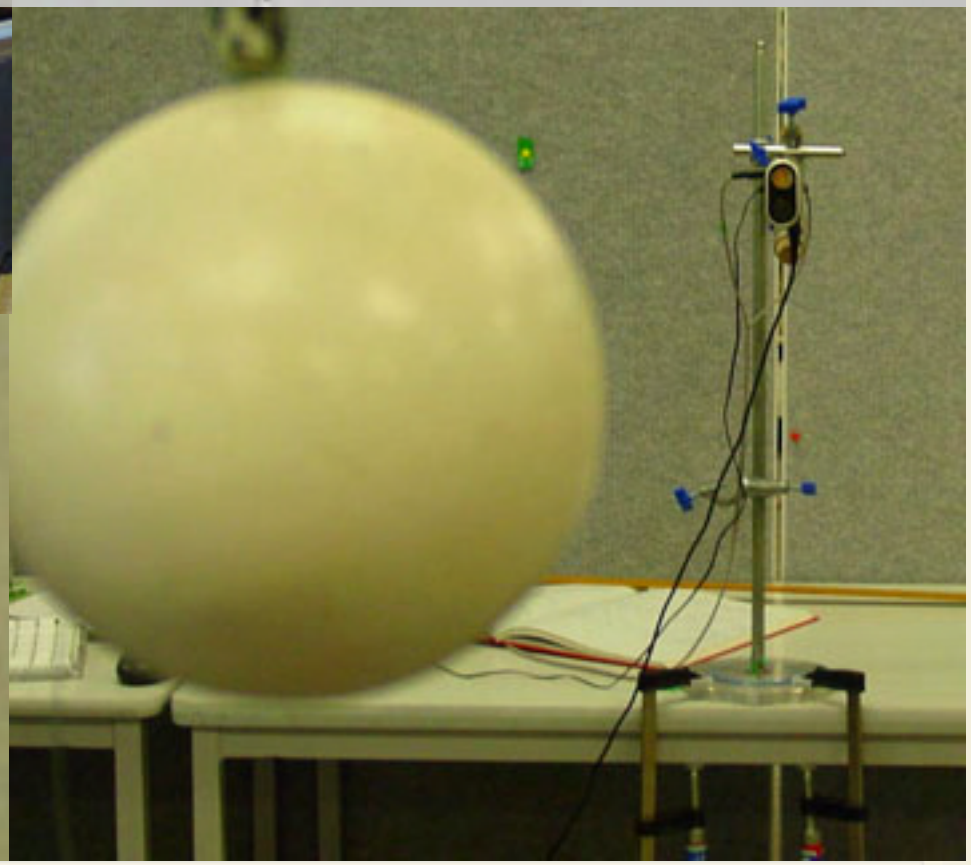
Experimental Setup

Control or measure every parameter

- ensonify head but not shoulders
- point beam at nose
- place head in known location
- eliminate noise
- electronic - ground
- acoustic - point into air
  - cross talk - objects
- aliasing - filter




LabVIEW File Edit View Project Operate Tools Window Help
Producer/Consumer Design Pattern (Data)
Time Domain Echo
Plot 0
Amplitude
Time
Power Spectral Density
Amplitude
fft bin
Start Scan
76
PSD error
0
Input size
1024
512
STOP
Elements in Queue
5
0



Classification

Mahalanobis distance between feature vectors

- linear classified that includes std
- Euclidean distance between normalised feature vectors
  - Features normalised by dividing mean by std
  - All features have an std of 1
  - Distance is measured in units of std

Feature Quality

Rank	Feature Name	Minimum Distance
1	Sum of Profile in Area C of face Model - Fig. 7.	0.0958276
2	Distance to Peak 1	0.0802839
3	Threshold 8	0.0638312
4	Length of Profile Area A to Area C – Fig. 9.	0.058637
5	Length of Profile to 75% Acoustic Area	0.0565977
6	Sum of Profile in Area A of signal features	0.0388602
7	Threshold 7	0.0343516
8	Sum of Profile in Area B of face Model – Fig. 7.	0.0296749
9	Length of Face Profile – Fig. 9.	0.0272006
10	Length of Profile Area B to Area C Midpoint	0.0242173

Minimum distance between feature values for 10 people

Results

Distance	1.46 to < 2σ	2 to < 3σ	3 to < 4σ	4 to < 6σ	> 6σ
count	5	9	12	9	10
% correct	53.58%	68.27%	86.64%	95.45%	99.73%

For 10 faces with 10 features

- Worst case - distance of 1.46σ -> 53.58% correct
- Only 10/45 distances > 6σ -> 99.73%

Conclusions

- Results poor
  - quality of features is poor
  - face recognition is a high-dimensional problem
  - effect of increasing range resolution
- mapped echo from face to depth of facial features
- will work for small groups of people
- may be possible to measure head tilt