

2007

Classifying still faces with ultrasonic sensing

P. J. McKerrow

University of Wollongong, phillip@uow.edu.au

K. K. Yoong

University of Wollongong

Publication Details

This article was originally published as McKerrow, PJ and Yoong, KK, Classifying still faces with ultrasonic sensing, Robotics and Autonomous Systems, 55, 702-710. Copyright Elsevier 2007.

Classifying still faces with ultrasonic sensing

Abstract

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 still face. Potential features for classification are identified using a facial model. The classification results for 10 faces encourage future research with a larger number of faces and with moving faces.

Disciplines

Physical Sciences and Mathematics

Publication Details

This article was originally published as McKerrow, PJ and Yoong, KK, Classifying still faces with ultrasonic sensing, *Robotics and Autonomous Systems*, 55, 702-710. Copyright Elsevier 2007.

Classifying still faces with ultrasonic sensing

Phillip M^cKerrow Kok Kai Yoong

School of Information Technology and Computer Science
University of Wollongong
Northfields Avenue
Wollongong NSW 2522, Australia
phillip@uow.edu.au

Keywords: face recognition, classification, CTFM ultrasonic sensing, echolocation

Abstract

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 still face. Potential features for classification are identified using a facial model. The classification results for 10 faces encourage future research with a larger number of faces and with moving faces.

1 Introduction

The shape of the human face is interesting for many reasons. Face shape is an important characteristic when recognizing one another. We enhance verbal communication with facial expressions produced by changing the shape of facial features. We describe faces with geometric descriptions both of the overall face and of individual parts. For example, we say some one has a moon shaped face.

The shape of our faces is similar, yet it varies from individual to individual. The shape of a face is symmetrical around a plane from front to back of the head, yet unsymmetrical around planes from side to side. Because of our ability to recognize one another by our faces, face recognition has become an important biometric for machine perception systems, commencing with the work of Bledsoe in the 1960s [5]. He calculated the distance between important facial features, such as the distance between the eyes, and used these as features in his face recognition algorithm.

Most research has focused on emulating face recognition by people with computer vision [13, 16]. Recognizing a face is a difficult task. Many features are required because faces are similar. Vision systems developed to recognize a few people appear not to scale up to many people. Also, they are often fooled by disguises [14].

Consequently, the quality of image based recognition systems installed at airports is considered to be poor. In 2002, the American Civil Liberties Union found that the Identix system installed at the Palm Beach International Airport failed to identify airport employees 53% of the time. A test at Logan International Airport in Boston on both Identix and Visage systems failed to identify people 38% of the time. The Cognetec system at Sydney International Airport performed better due to better lighting control and passenger cooperation. In the face recognition vendor test (FRVT) conducted in 2002, none of the 10 systems tested achieved a satisfactory level of performance [6].

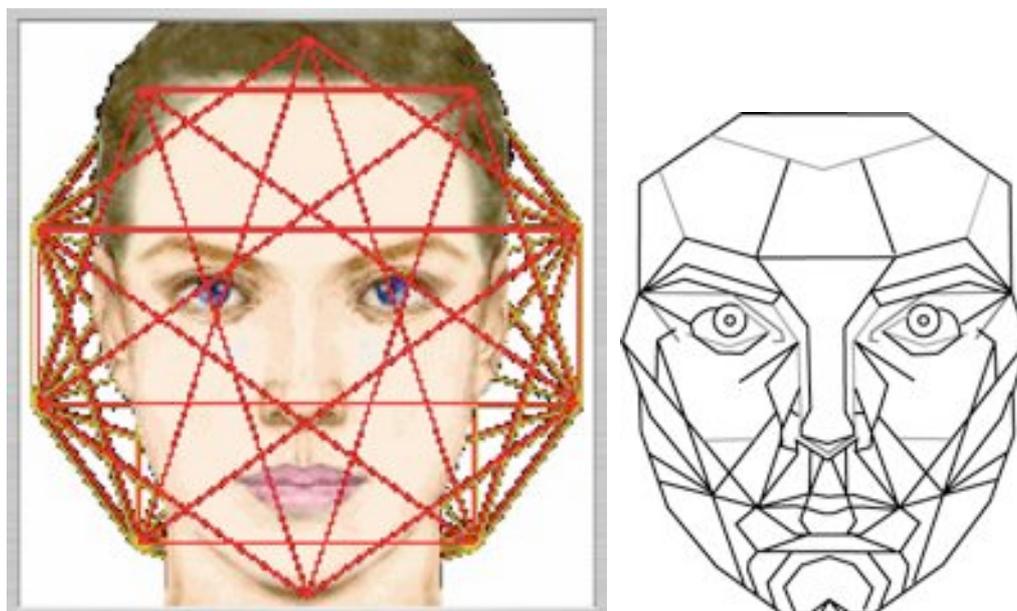
The problem we wish to address in this paper is: is it possible to recognize or classify a human face with the information in the echo of an ultrasonic chirp from that face? This question has several related questions. Is the geometry of faces sufficiently different to produce measurable differences in the echo? Can a face be recognized by geometry alone? What is the effect of face geometry on the echo?

Dror *et al.* [3] trained a neural network to recognize 5 faces using the time-frequency spectrogram of the echo as input. The network recognized the 5 faces with an accuracy of 96%. When a sixth face was added the accuracy dropped to 81% and a seventh face could not be recognized. Thus, their system did not scale.

While other research into ultrasonic sensing can point to bats as a working model, for example sensing roughness [9], we have seen no paper that suggests that bats can recognize human or even bat faces. The motives for doing this research are interest: to see if it is possible, and the outcome of previous research: that ultrasonic sensing is a geometry sensor that measures range to and near orthogonal area of surfaces.

A previous paper [15] presented initial ideas and a case study comparing a cardboard, a polystyrene and a human face. This paper develops a robust measurement method from those ideas and reports on the classification of ten human faces. First, we look for geometric models of faces. Second, we look at the relationship between the echo from the front of a face and the physical features of the face. From this we predict

possible features for classification. We discuss the difficulty of setting up an experimental approach that will guarantee accurate and repeatable results. Finally, we discuss the classification of 10 faces.



a. part of grid overlaid on face

b. dividing face into regions of similar depth

Figure 1. Marquart Beauty Analysis repose mask front view

Face Geometry

A common approach to modeling an object in computer graphics is with a set of connected facets. The flat facets are at different angles and distances from the camera, and connect to produce a continuous surface that looks like a solid object.

Facet models are regularly used in ultrasonic sensing. For example, when classifying rough surfaces the surface can be modeled as a set of small facets. Combining a facet model of the rough surface with a model of ultrasonic sensing produced excellent classification results [9].

The echo from a chirp of ultrasonic energy impinging on an object is the superposition of echoes from those facets whose surfaces are near orthogonal to the incident beam. Thus, the information in the echo is determined by the geometry of the ensonified object. The orientation of a facet relative to the incident beam determines the direction that the echo from that facet travels.

Most human made surfaces reflect ultrasonic energy specularly. Consequently, only those facets that are near orthogonal in orientation to the incident chirp reflect energy to the receiving transducer. Smooth surfaces are those with a depth of texture much less than a quarter wave length. As the wavelength at 100KHz in a standard atmosphere at 20°C is 3.434 mm, the surfaces of a young person's face are relatively smooth. By comparison, as a person ages their face wrinkles and becomes rougher. Thus, the reflection of ultrasonic energy from a face will be different to the reflectance fields measured for light [2].

The distance to a facet from the transducer determines the time at which reflection occurs relative to the reflections from other facets. The amount of energy reflected to the receiver is a function of the orientation, roughness and impedance of the facet. As most surfaces have a high impedance relative to air, almost 100% of the energy is reflected. Only materials that are mainly air, like foam, absorb a measurable amount of energy. Thus, we only expect surface reflections and no reflections from epidermal and dermal layers that occur with light [2].

The result is that the echo contains information about the geometry of the object: time is proportional to range (facet depth) and amplitude is proportional to reflected energy. The reflected energy is a function of the area of the facet projected onto a surface normal to the axis of the impinging beam.

We are interested in whether the geometry of human faces is sufficiently different to produce measurable differences in their echoes that we can use to distinguish between faces. So we are interested in models of face geometry that match the facet model of objects that we described above. We may have looked in the wrong place but we found very little material on face geometry [7] that models faces as facets. Most papers model faces at either a much finer level for realistic visual reproduction or are anatomy based [12] for animation.

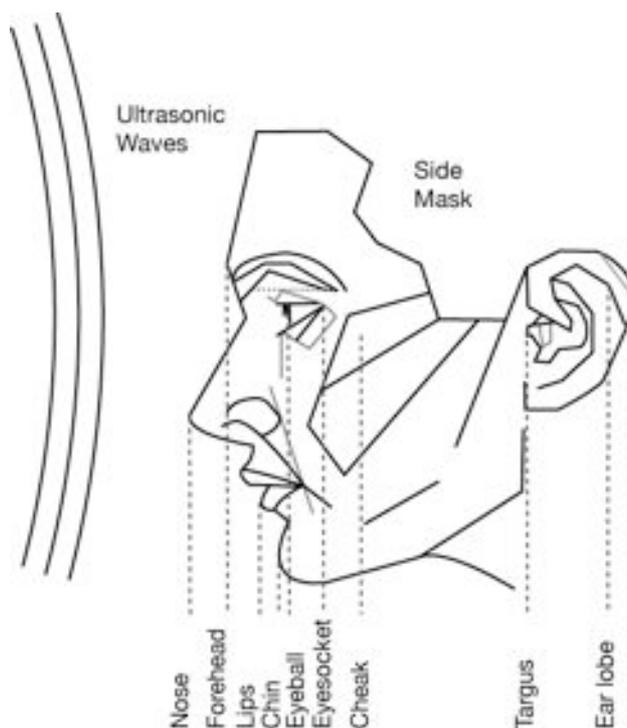


Figure 2. Side view of repose mask showing relationship to impinging ultrasonic waves.

A model that we found helpful for thinking about face shape is the “Repose Mask” model developed for analyzing the beauty of a face [11]. It is a universal face model that models the face with a set of polygonal facets. The mask in Fig.1. divides the face into geometric regions of similar depth. Marquart use this mask as a universal model of facial features. They discuss how people from different ethnic background differ from the model. For example, Asian people tend to have wider faces than shown by the mask, Europeans have wider noses and Africans have thicker lips.

The side view is useful when thinking about ultrasonic sensing because it shows the relative depth of the facial features (Fig. 2.). When a face is vertical, we can expect the first echo to be from the tip of the nose, and the last echo to be from the ear lobe. Between these, various facial regions will produce echoes whose amplitude is proportional to the area of the face that is nearly orthogonal to the beam. The front view of the mask indicates that the forehead is one such region. Also, we expect corner reflectors, like eye sockets, to produce significant echoes. The distance between facial features should be measurable from the echo.

CTFM Ultrasonic Sensing

In this research we used the K-sonar CTFM (Continuously Transmitted Frequency Modulated) sensor developed by BAT [1] as a mobility aid for blind people (Fig. 3.). One transducer is used for transmission and one for reception. A single 19mm diameter transducer has a theoretical beam angle of 19.32° from axis to first minima (Fig. 4.). Combining two transducers to form a transmitter and receiver, the vertical diameter is 47mm and the theoretical horizontal beam angle is 7.6° .



Figure 3. K-Sonar ultrasonic mobility aid - sits on a white cane.

Classifying still faces with ultrasonic sensing

We set the CTFM system to transmit a downward swept sine wave (f_{sweep} is 100kHz to 50kHz) every 100msec (sweep period t_s). The ultrasound energy reflects from objects and returns to the receiver as an echo. The echo is a delayed and filtered version of the transmitted signal. A demodulation sweep, derived from the transmitted sweep, is multiplied with the received echo in the time domain. The outputs of this multiplication are sum and difference frequencies (Fig. 5).

The distance of flight information is contained in the difference frequencies (f_a is 0 to 5kHz), where frequency is proportional to range (Fig. 5.) and amplitude is proportional to surface area. This time domain signal is converted to a power spectrum with an FFT to give a range-energy echo (Fig. 6.). The amplitude in frequency-bin i is the energy reflected from surfaces in a spherical annulus at range r_i (Fig. 4.).

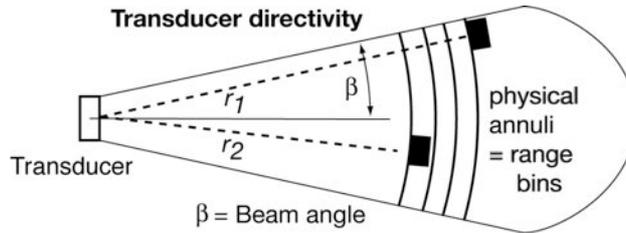


Figure 4. An ultrasonic transducer emits a beam of energy. R = range to orthogonal surface.

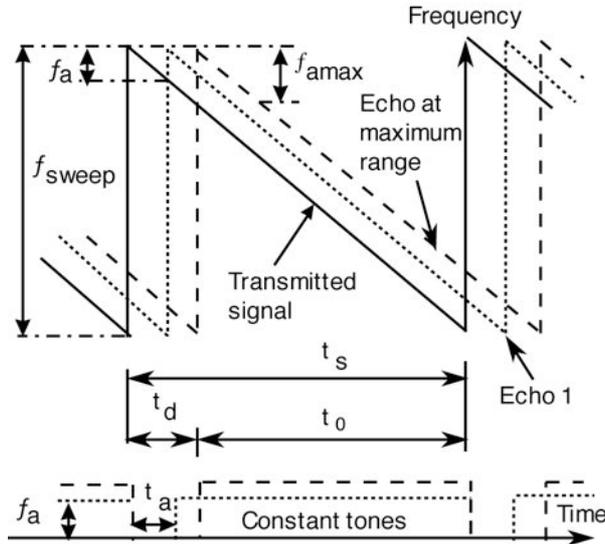


Figure 5. CTFM demodulation – multiplying the echo by the transmitted signal produces a set of different tones where frequency is proportional to range to object.

The audio tone for a surface is continuous from the time at which the echo arrives until the end of the sweep. Thus, the sweep consists of two time periods: the time to the arrival of an echo from maximum range (t_a) and the time to capture the samples (t_0) for the FFT. In the time domain, the complexity of the audio signal is proportional to the geometric complexity of the target. It contains a tone for each surface where sound is scattered back to the receiver. The purpose of calculating the power spectrum is to separate these tones.

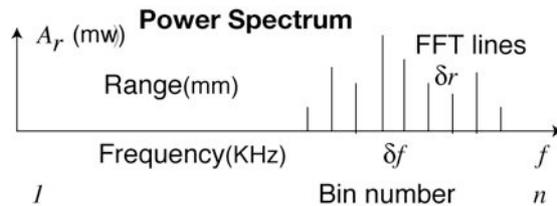


Figure 6. Power spectrum of echo – frequency (bin number) is proportional to range and amplitude to echo energy at that range.

A 1024 point FFT divides the echo into $n = 512$ frequency bins of width $\delta f = 10\text{Hz}$. The bins represent a set of concentric annuli, each $\delta r = 3.867\text{ mm}$ thick (Fig. 4.). This dimension is the range resolution of the system. It can be changed by changing the period of the sweep or the environment conditions. We measure it when setting up an experiment to compensate for minor differences between the timing of the sweep period in different sensors as well as air temperature.

The frequency of an FFT line f_i is proportional to the range r_i to the annulus containing the surfaces that produced that component of the echo. The amplitude A_i of the FFT line at range r_i is the absolute value of the complex number output from the FFT. It is proportional to the pressure of the echo and hence to the area of the surfaces normal to the receiver in the annulus.

Relationship between echo and face depth

The side view of the repose mask in Fig. 2. shows that facial features are separated in depth. Thus ultrasonic energy impinging on those features will produce echoes at different times. These echoes should show up in the power spectrum in different bins. So we decided to conduct an experiment to test this idea.

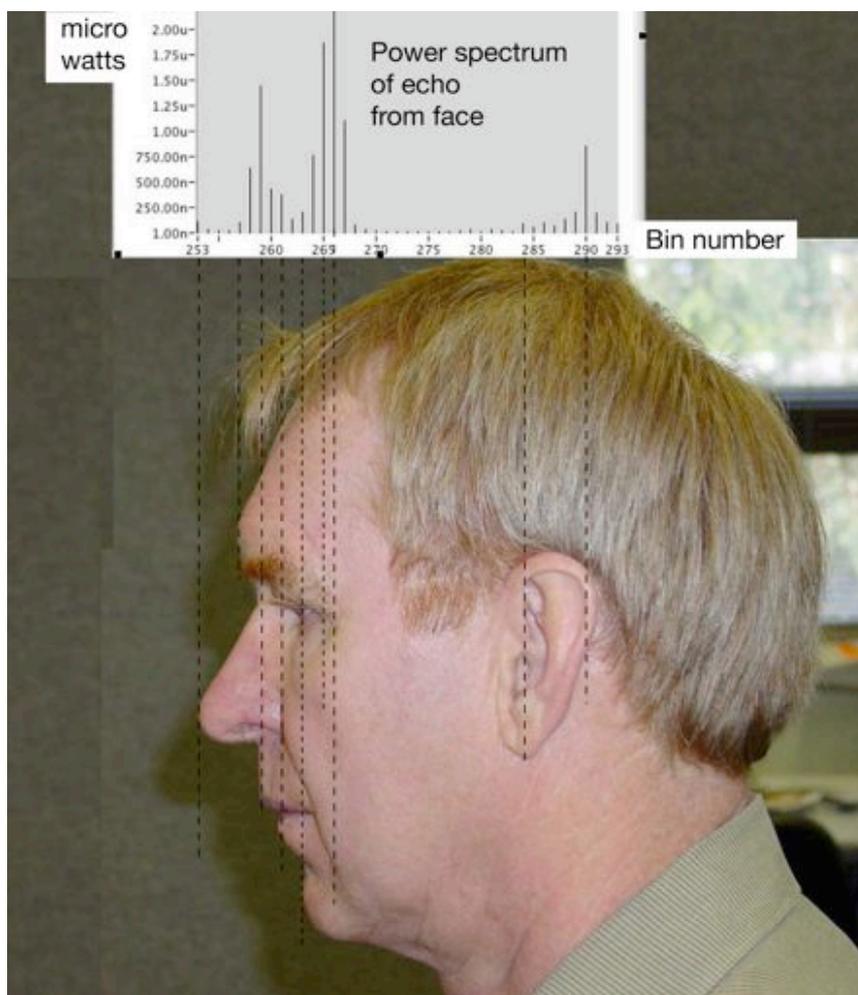


Figure 7. Power spectrum of echo from author's face lined up with image of the side of his face.

Fig. 7. Shows the power spectrum of an echo from the author's face lined up with a photograph of the side of his face. We measured the millimeters per bin of the ultrasonic echo and scaled the photograph to the same scale. The echo first exceeded the noise removal threshold at bin 253, so we lined this bin up with the tip of his nose. As shown in Fig. 7., the echo lines up very well with forward oriented facial features that we would expect to return an echo to the receiver.

The first peak lines up with the section of his face above his lips and the bridge of his nose. The second peak lines up with the outside of his eye sockets. The last peak lines up with the forward facing part of his earlobe. This result shows that the echo consists of echoes from facial features and indicates the possibility of using distances between peaks as a feature for classification.

Next, we measured ten people and performed a similar match. For each person we determined the bin numbers that correspond to specific facial features. We gathered these together in a table that we call an Acoustic Face-Geometry Table (Table 1.). The underlined bin number is the bin with the highest amplitude for each person. Where two bins are underlined the peak occurs between them.

In general, the most energy is reflected from the depth of a person's eye. The reflector could be their eye socket or their cheeks. For person 5, the peak occurs earlier in the region of their lips. From the photos, this person has a flat face. The depth of her face from top of lips to edge of eye socket is two thirds of that of the author (Fig. 7) while the width is the same. A minor tilt of her head would line her forehead up with the top of her lips creating a large echo at that range.

From the photographs, we saw that both person 5 and person 9 have short noses. Person 7 also has a short nose, but his face is tilted up in the photo so that his nose and lips line up. Person 4 has her face tilted down. These results demonstrate that the relative depth of a person's facial features is sensitive to the tilt of their head.

The photographs were taken after the echoes were recorded, so the person may have changed the tilt of their head. Obtaining accurate photographs at the same time as recording the echoes, with a ruler included for scaling, is the one part of the experimental procedure for which we are still to find a robust method.

Shooting video during the experiments and recording the echoes on the sound track may help. Also, this would enable experimentation with hybrid vision/sonar sensing. Image data is not used in classification but rather for understanding.

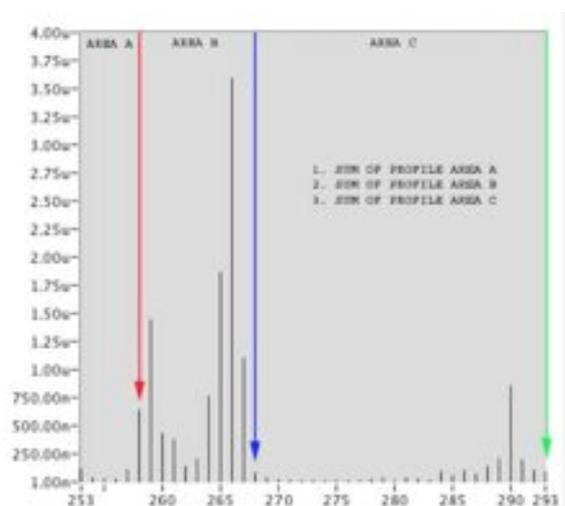


Figure 8. Dividing acoustic density profile up into 3 regions based on the face model.

Features

Having demonstrated an alignment between the output of the power spectrum of the echo and the geometry of a person's face, the next step is to identify features in the power spectrum that may be due to geometric features in the face. In previous research into the classification of plants from their echo [8] and the classification of rough surfaces from their echoes [9] we developed sets of features. In each case, we were able to give a geometric interpretation to those features, but like most inverse problems many echo features don't have a one to one correspondence with geometric features.

As in the previous research, the length of the window (called acoustic density profile) is an important feature and in this case we suggest that it is the depth of the face from nose to ear lobe. A second feature of importance is the distance to the highest peak (peak 1). In the acoustic face-geometry table, we saw that this feature often corresponds to the distance from the tip of the nose to the edge of the eye. When it doesn't, it indicates that the person has a large flat region on their face.

To generate a set of features that may relate to geometry we divided the acoustic density profile into 3 regions, based on the observation that there are three distinct peaks in the profile in Fig. 7. Two approaches to dividing the face profile are illustrated in Figs. 8 and 9. The first approach (Fig. 8.) is based on the matching of facial features in the model to the acoustic density profile. Area A includes the facial features from the nose to the chin. Based on the average nose to chin distance of the 10 people in Table 1, it is 6 bins long. Area B is from the end of area A to the the mid point of the face. The mid point is the average of the nose to ear distance for the 10 people and is at bin 15. Area C is from bin 16 to the end of the ear.

Three features are shown in Fig. 8. The sum of the profile in each area is a measure of the energy reflected in that area and is proportional to the area of the surfaces that reflect energy toward the receiver. An additional feature of this

type, that has proved useful in previous research, is the length to 75% of the sum of the acoustic density profile. This feature probably includes the front and mid portions of the face but not the ears.

The second approach to dividing the profile into three regions is to select regions around the three peaks (Fig. 9.). In Fig. 7. the first peak lines up with the flat part of the face between the upper lip and the nose and with the bridge of the nose. The second peak lines up with the edge of the eye and the third with the front of the ear lobe. Thus, this approach is more related to the characteristics of the signal than to the model. Fig. 9. Shows 14 features that can be obtained using this approach. A number of other features were also considered.

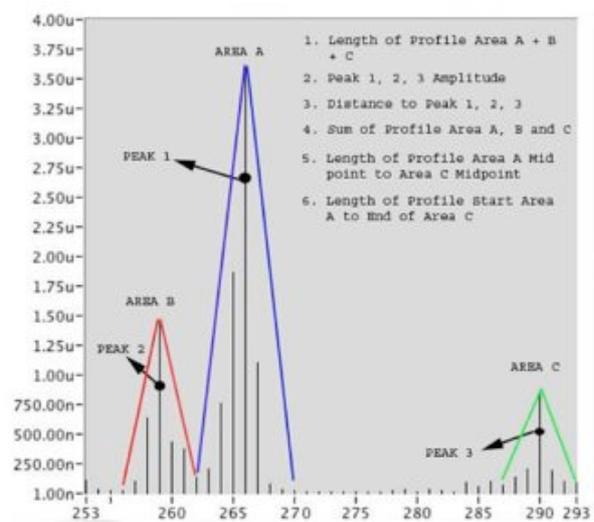


Figure 9. Dividing acoustic density profile up into 3 regions based on signal features.

Experiment setup

The above model and feature set determine what we want to measure when sensing faces. Building an experimental setup where we could control or measure every parameter was a challenge. We wanted to get the best echoes possible so that we could determine the potential of ultrasonic sensing for recognising faces.

The transmitter produces a beam (Fig. 4.). Our aim was to have the person's nose at the centre of the beam with all their head ensonified by the beam, but not their shoulders. This involved three issues: aiming the sensor, placing the person's head at a known location and removing all sources of noise.

To ensure the face was within the beam, we measured several faces to find that the widest face was 204mm and the longest face was 258mm. To allow for some error in face position and for face movement we setup the beam to ensonify a larger area at the range of the nose. We measured the beam angle to be $\pm 11^\circ$ vertical and $\pm 10^\circ$ horizontal. At 1m these angles give an ensonified area of $324 * 363$ mm (w*h).

To aim the sensor we hung a 100mm diameter sphere from the roof at 1m from the transducer and at the head height of a seated person. Then we panned and tilted the sensor until the echo amplitude was maximum. At these orientations, the sensor should be pointing at the centre of the sphere. To place the person's head, we seated them on the side of the sphere away from the sensor and adjusted the height of the chair until their nose was at the same height as the centre of the sphere and just touching the sphere. We found it was important to get them to sit up straight and point their nose toward the sensor, because the relative depth of facial features varies with head orientation.

Then we swung the sphere out of the way, so that we could ensonify their face and capture sets of 64 echoes in various poses. In the following processing, each echo is windowed to remove the bins before and after the face. The threshold for windowing is set at the mean plus five standard deviations of the measured noise (25 nanoWatts). The noise level is measured by pointing the sensor into air with no objects in it.

The third issue that has to be dealt with is noise. Noise can arise from several sources: electronic, acoustic and signal processing. To measure the electronic noise, we disconnected the receiver from the analog to digital converter (ADC), shorted the input to ground and measured what should be the minimum noise. Also, this measurement should provide a reference point for the following measurements.

To measure the acoustic noise, we reconnected the receiver and removed all objects from in front of the sensor so that it should be pointing into air. Any echoes from objects should show up and we remove the object. Also, this measurement is a measurement of total system noise that can be used to calculate thresholds [8, 9].

At this point we found we had no cross talk from transmitter to receiver but we had signal processing noise. As we were expecting demodulation tones up to 5KHz, we set the ADC sample rate at 12KHz. Any signal in the echo with a frequency above 6KHz will alias down into the range of interest. We found that we had several of these.

The audio output of the mobility aid is produced by a digital to analog converter and amplified to drive a set of head phones. It has no filtering at all on it. It relies on the head phones to remove all high frequencies before they reach the

listener's ear. So we had both electronic noise from the conversion process and signals from objects that were further away than the maximum range for the current settings. We removed both of these with a low-pass filter. We use ac coupling of the echo to the ADC to remove dc offset, so we effectively have a band-pass filter.

We are confident that the resulting experimental setup is free from measurement error. It enabled us to quickly record several sets of echoes from each person, with all variables under control. In this paper, we only discuss results for echoes where the person's face is pointing toward the sensor and is still. We also recorded data sets for facial expressions and head motion.

Unfortunately, the results that we published in a previous paper [10] were poor due to noise causing the windowing algorithm to occasionally return the wrong window. The current windowing algorithm uses 3 settings to find the window: a threshold, a start bin, and a check range. It ignores all fft bins prior to the start bin and then looks for the first bin above the threshold. Having found the start of an echo it searches until it finds a sequential group of bins (check range) less than the threshold to confirm that it has found the end of the echo.

Setting the window threshold at the mean of the noise plus 5 standard deviations [9] assumes Gaussian noise. While this appears to be valid for the electronic and environmental noise, it is not valid for low energy reflections from side lobes or the aliasing of distant objects. These reflections appear in the signal prior to the face echo as low energy spikes (typically less than twice the threshold). The purpose of the start bin setting is to ignore the bins that are known to be prior to where we expect the face echo to be.

As the start distance was set too low, we occasionally windowed a noise spike instead of a face echo resulting in the calculation of incorrect features values in later processing. As we know the geometry of the experimental setup, we were able to increase this start distance to remove the false windows. The result is a significant improvement in both feature quality and classification results compared to those we published previously [10].

Classification

Classification is the process of giving meaning to a set of observations (echoes) by mapping a multidimensional feature vector to one of a set of possible meanings (faces). To evaluate the quality of features and to classify faces we use the Mahalanobis distance between feature vectors [9]. A feature vector is obtained for each face by calculating the mean and standard deviation of the features extracted from the echoes of it. Then we calculate the distance between vectors of these features to determine their ability to discriminate faces. Thus, the resultant measures indicate the likelihood of each of the echoes identifying one of the faces and not another.

While a non-linear classifier, such as a neural network, may give better classification results [4], a statistical classifier can be used to measure the quality of the classification. To make the calculations robust, and the results easy to compare, the Mahalanobis distance is calculated by calculating the Euclidean distance between the normalized feature vectors. Features are normalized by dividing their means (μ) by their standard deviations (σ). This transforms all features into a space where they have a standard deviation of 1 and all distances are measured in units of standard deviation.

We use two measures to examine the quality of a feature: the smallest distance between any two values of the feature and the count of distances between feature values where the distance is less than a threshold. We take the values for a feature for all the objects being measured, in this case 10 people. These form a set of points in 1D space, where the closest two points are the least separable.

Table 2 shows the best 10 features based on the minimum distance between feature values for the 10 people for each feature. Four threshold features are included. They are multiples of the noise threshold ($\mu+5\sigma$ of measured noise = $25n_w$) as used in plant recognition [8]. The rest are features based on face geometry. There are more features from the signal model (Fig. 9.) than from the average face model (Fig. 7.). This is encouraging as it reflects the different geometry of the faces. In fact, the best feature maps to the distance from the edge of the eye to the front of the ear lobe. The fact that the distances are significantly less than 1σ indicates that there are similar faces in the set and classification will be hard.

The final step is to use these features to classify the 10 faces. Table 3 shows the results when using the best three geometric features in Table 2 (Features 21, 6 and 1). 35 of the 45 distances are better than 6σ . Of the 10 distances less than 6σ , only 2 are less than 3σ . We divide the distance (in standard deviations) between face vectors by 2 to get the percentage of correct classifications for each face when comparing them. When the distance is 6σ between two faces, 99.73% of the echoes are within 3σ of the reference vector for a face. The worst case in Table 3 is a distance of 2.05σ , which means the classification is correct 68.3% of the time when comparing Tiffany's face (female) to Wei's face (male). When the 4th geometric feature (Feature 13) is added to the vector, only 3 distances are less than 6σ (3.2σ , 4.8σ and 5.94σ).

Results and conclusions

We are interested in whether the geometry of human faces is sufficiently different to produce measurable differences in their echoes that we can use to distinguish between faces. The results of this research demonstrate that such

discrimination is achievable for a small set of faces. Hence, a robot working with a few people could use this sensing to identify its co-workers. However, these results may not scale up to recognizing a large number of people.

In comparison to the results that were achieved when classifying rough surfaces [9], these results are not as good. When classifying surfaces based on roughness, a moving sensor achieved a classification better than 99.73% for 12 surfaces using 5 features. Here, a still sensor achieved a classification better than 99.73% for 10 faces using the best 6 features in Table 2 for all pairs of faces except 1. However, these results are better than those previously reported for face classification with ultrasonic sensing.

This research did not look at the effect of head tilt, head rotation or changing facial expressions on the quality of the classification. We have observed sensitivity to tilt. Tilt changes the relative spatial distance between facial features. So it may be possible to develop a sensor to measure the orientation of a person's head. We got people to turn their heads to various angles and to tilt them up and down while we recorded the echoes from their faces. We also recorded their faces with various facial expressions. However, we are yet to examine that data.

This research has opened up a number of questions to be pursued in future research. Each of the feature extractors should be examined again to see if it can be improved. Also, the windowing algorithm need further research to handle the different types of noise found in different sensing situations. The mapping from face geometry to echo should be studied in more detail to see if there are other features in the data and to achieve better discrimination of facial features. Another area of research is whether higher range resolution would improve the separation between facial features and make them easier to find in the echo.

This research has shown that the echo from a face can be mapped to the depth of facial features. We found that the features that map to the geometry give good classification with a small number of faces (10). Exploring whether this scales up to a large number of faces and the impact of face motion on the classification are topics for future research. From the geometric similarity of faces that we have observed, we conclude that face classification is a high-dimensional problem that requires a lot of features.

References

- [1] BAT, K-sonar, 2006. <http://www.batforblind.co.nz/>
- [2] Debevec, P., Hawkins, T., Tchou, C., Duiker, H-P., Sarokin, W. and Sagar, M., 2000. Acquiring the Reflectance Field of a Human Face, Proceedings of SIGGRAPH 2000, pp. 145-156, July, ACM Press
- [3] Dror IE., Florer FL., Rios D. and Zagaeski M. 1996. 'Using artificial bat sonar neural networks for complex pattern recognition: Recognizing faces and the speed of a moving target'. *Biological Cybernetics*. Vol. 74, pp. 331-338.
- [4] Harper, N.L. and M^cKerrow, P.J. 1997. 'Recognition of Plants with CTFM Ultrasonic Range Data using a Neural Network', Proceedings, IEEE International Conference on Robotics and Automation, Albuquerque, ICRA'97, April, pp 3244-3249.
- [5] Kelly G. 2004. 'The Past Perfect Promise of Facial Recognition Technology'. *ACDIS Occasional Paper*. : <http://www.acdis.uiuc.edu/Research/OPs/Gates/contents/part2.html>.
- [6] Kung SY., Mak MW and Lin SH. 2005, 'Biometric Authentication By Face', *Biometric Authentication*. Upper Saddle River, NJ, Prentice Hall, pp. 241-276.
- [7] Li, S.Z. and Jain, A.K. 2005. Handbook of face recognition, Springer.
- [8] M^cKerrow, P.J. and Harper, N.L. 2001. 'Plant acoustic density profile model of CTFM ultrasonic sensing', *IEEE Sensors Journal*. Vol 1, No 4, 2001, pp 245-255.
- [9] M^cKerrow PJ. and Kristiansen BE. 2006. 'Classifying surface roughness with CTFM ultrasonic sensing', *IEEE Sensors Journal*, Vol. 6. No. 5., October, pp 1267-1279.
- [10] McKerrow, P. J. & Yoong, K. 2006, 'Face classification with ultrasonic sensing', *Towards Autonomous Robotic Systems 2006 (TAROS 2006)*, Department of Computing, Imperial College, London, pp. 111-117.
- [11] Marquardt Beauty Analysis. 2005. 'Normal Face Variations- From The Mask': http://www.beautyanalysis.com/index2_mba.htm
- [12] Romdhani, S., Ho, J., Vetter, T. and Kreigman, D.J. 2006. 'Face Recognition Using 3-D Models: Pose and Illumination', Proceedings of the IEEE, Vol 94, No 11, Nov, pp 1977 – 1999.
- [13] Shinha, P., Balas, B., Ostrovsky, Y. and Russell, R. 2006. 'Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About', Proceedings of the IEEE, Vol 94, No 11, Nov, pp 1948 – 1962.
- [14] Yang, M.H., Kreigman, D.J. and Ahuja, N. 2002. 'Detecting Faces in Images: A Survey', *IEEE Trans on PAMI*, Vol 24, No 1, Jan, pp 34 – 58.

Classifying still faces with ultrasonic sensing

- [15] Yoong, K.K. and McKerrow, P.J. 2005. Face recognition with CTFM sonar, Proceedings of the 2005 Australasian Conference on Robotics & Automation, Sydney, December, 10 pages.
- [16] Zhao W., Chellappa R., Phillips P.J. and Rosenfeld A. 2003. 'Face Recognition: A Literature Survey'. *ACM Computing Surveys (CSUR)*. Vol. 34, no. 4, pp. 399-458.

Table 1. Acoustic Face-Geometry Table showing the relationship between facial features and echo bins for 5 females and 5 males (bottom half), with the bin where peak amplitude occurred underlined.

Person	Nose	Fore head	Lips	Chin	Eye	End of Eye	Start of Ear	Ear Lobe	End of Ear
Helen	243	244	246	248	<u>248</u>	251	269	271	268
Lily	254	259	256	<u>260</u>	<u>260</u>	264	283	288	288
Pansy	247	251	248	252	<u>253</u>	<u>257</u>	273	276	283
Tiffany	250	252	254	259	<u>256</u>	<u>259</u>	277	282	285
Wei	259	263	<u>262</u>	264	264	267	284	286	292
Adrian	255	260	259	262	<u>263</u>	265	283	285	295
Kwan	254	258	256	255	<u>260</u>	263	281	282	292
Bill	252	256	255	258	<u>258</u>	<u>261</u>	279	280	290
Daniel	259	266	262	262	<u>266</u>	270	290	292	299
Phillip	253	257	259	263	263	<u>265</u>	282	284	291

Table 2. Best 10 features ranked by maximum of the minimum distance between feature values

Feature	Feature Name	Minimum Distance (σ)
21	Distance Peak 1 to Peak 3 - Fig. 9.	0.272483
25	Count above threshold 4	0.192743
1	Length of Face Profile	0.144017
6	Sum of Profile in Area C – Fig. 7.	0.112629
22	Count above threshold 1	0.0954724
13	Distance to peak 3 – Fig.9.	0.0907741
26	Count above threshold 5	0.0862213
3	Average acoustic area	0.0674363
10	Peak 2 amplitude – Fig. 9.	0.0557841
30	Count above threshold 9	0.0552358

Classifying still faces with ultrasonic sensing

Table 3. Classification results for 10 faces with the 3 best geometric features in Table 2: Features 21, 1 and 6. 10 of the 45 distances between the feature vectors are less than 6σ (underlined) and 2 of those are less than 3σ .

From Face y	To Face y+1	Face y+2	Face y+3	Face y+4	Face y+5	Face y+6	Face y+7	Face y+8	Face y+9
1. Helen	<u>5.94</u>	11.93	6.58	6.26	23.41	<u>4.55</u>	41.85	43.59	<u>5.83</u>
2. Lily	13.74	6.02	7.22	24.44	<u>2.75</u>	45.54	45.16	7.9	na
3. Pansy	12.81	13.45	11.78	11.13	30.57	32.64	8.62	na	
4. Tiffany	<u>2.05</u>	24.16	<u>5.06</u>	43.28	45.39	<u>4.61</u>	na		
5. Wei	25.04	6.22	43.94	46.09	<u>4.85</u>	na			
6. Adran	22.14	19.40	21.5	20.3	na				
7. Kwan	41.22	42.95	<u>5.44</u>	na					
8. Bill	<u>4.04</u>	39.11	na						
9. Daniel	41.25	na							
10. Phillip	na								