A performance adequate computational model for auditory localization

Wing Chung
Systems Engineering and Design Automation Laboratory, Department of Electrical and Information Engineering, University of Sydney, Sydney 2006, Australia

Simon Carlile[a)]
Auditory Neuroscience Laboratory, Department of Physiology, University of Sydney, Sydney 2006, Australia

Philip Leong
Department of Computer Science and Engineering, The Chinese University of Hong Kong,
Shatin NT, Hong Kong

(Received 13 February 1998; revised 26 April 1999; accepted 23 July 1999)

A computational model of auditory localization resulting in performance similar to humans is reported. The model incorporates both the monaural and binaural cues available to a human for sound localization. Essential elements used in the simulation of the processes of auditory cue generation and encoding by the nervous system include measured head-related transfer functions (HRTFs), minimum audible field (MAF), and the Patterson–Holdsworth cochlear model. A two-layer feed-forward back-propagation artificial neural network (ANN) was trained to transform the localization cues to a two-dimensional map that gives the direction of the sound source. The model results were compared with (i) the localization performance of the human listener who provided the HRTFs for the model and (ii) the localization performance of a group of 19 other human listeners. The localization accuracy and front–back confusion error rates exhibited by the model were similar to both the single listener and the group results. This suggests that the simulation of the cue generation and extraction processes as well as the model parameters were reasonable approximations to the overall biological processes. The amplitude resolution of the monaural spectral cues was varied and the influence on the model's performance was determined. The model with 128 cochlear channels required an amplitude resolution of approximately 20 discrete levels for encoding the spectral cue to deliver similar localization performance to the group of human listeners. © *2000 Acoustical Society of America.* [S0001-4966(99)04411-2]

PACS numbers: 43.64.Bt, 43.64.Ha, 43.66.Qp [RDF]

INTRODUCTION

Humans can locate the source of a sound with remarkable accuracy using a variety of acoustic cues (Carlile, 1996). The location-dependent information contained in the sounds at each ear results from the interaction between the auditory periphery and the incident sound. The binaural localization cues include the interaural time difference cue (ITD) and the interaural level difference cue (ILD) (Middlebrooks and Green, 1991). The ITD operates principally at low frequencies and, conversely, the ILD is a reliable localization cue for the middle to high frequencies. Because of the relative symmetry of the ears on the head, a set of points in space can have the same binaural time or level values. That is, a binaural cue defines a ''cone of confusion'' centered on the interaural axis which leads to ambiguities in the vertical position of the sound source and front–back confusions (Oldfield and Parker, 1986). The auditory system most probably uses the spectral cues provided by the location-dependent filtering of the outer ear to resolve the cone of confusion (Middlebrooks, 1992; Carlile, 1996).

The head-related transfer function (HRTF) is defined as the acoustic transformation function from a point in space to the outer ear and describes the location-dependent filtering of a sound by the auditory periphery. The HRTF captures both the frequency domain and time domain aspects of the cues to a sound's location. Various models of the peripheral processing by the auditory system suggest that the fidelity of the acoustical information encoded by the nervous system is considerably degraded in the frequency domain when compared to the fidelity with which the HRTF is routinely measured (see Carlile and Pralong, 1994).

In the work reported here, we were interested in developing a model of localization that combined biologically plausible processing of the acoustical input with the input–output mapping provided by an artificial neural network (ANN). There were several key motivators for this approach.

First, preprocessing the input to the ANN in a biologically plausible manner would ensure that the mapping provided by the ANN would be a more reasonable model, in performance terms, of human localization performance. Thus, our first objective was to develop a model with a number of biologically plausible constraints that would provide a similar performance level as that found in humans. This necessitated the degradation of the model over the best that could be achieved without these constraints.

Second, a model with human-like performance could

[a)]Electronic mail: simon@physiol.usyd.edu.au

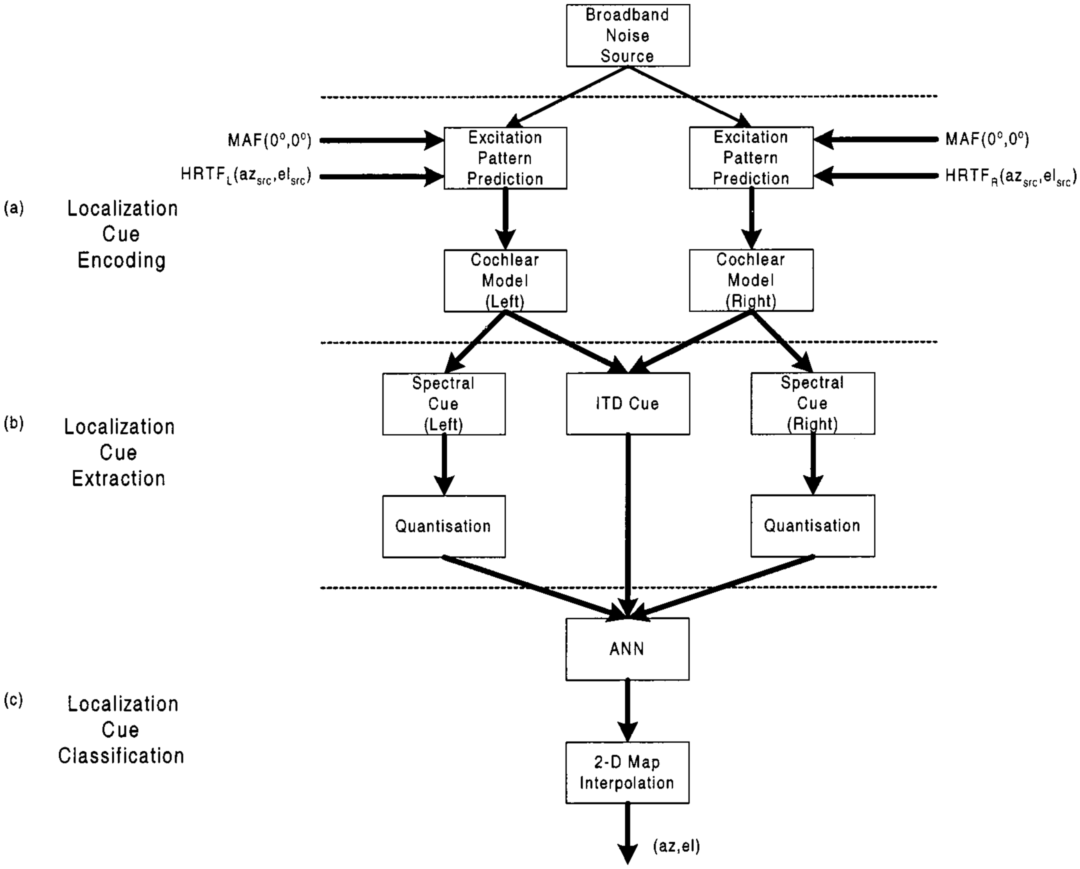
Broadband Noise Source
MAF(0°,0°)
HRTF$_L$(az$_{src}$,el$_{src}$)
Excitation Pattern Prediction
Excitation Pattern Prediction
MAF(0°,0°)
HRTF$_R$(az$_{src}$,el$_{src}$)
(a) Localization Cue Encoding
Cochlear Model (Left)
Cochlear Model (Right)
Spectral Cue (Left)
ITD Cue
Spectral Cue (Right)
(b) Localization Cue Extraction
Quantisation
Quantisation
ANN
(c) Localization Cue Classification
2-D Map Interpolation
(az,el)

FIG. 1. Localization model architecture.

then provide the basis for exploring aspects of preprocessing to the ANN by varying biologically constrained parameters and observing the impact on the subsequent localization performance. In this way we could explore the limits of the ‘‘biological resolution’’ of inputs to the model that are necessary to sustain human levels of localization performance. In addition, benchmarking the model’s performance against human localization performance would provide some insights into the likely biological relevance of various encoding and output parameters of the model.

Third, the output of the ANN was postprocessed using a spatial interpolator that attempted to model the behavior of a neuronal population of spatial location detectors. Previous neurophysiological studies of auditory space maps in the deep layers of the Superior Colliculus of mammals and the MLD of Owls have demonstrated significant difference in the size of the auditory spatial receptive fields of neurones in these nuclei [see King (1994) for review]. It has been proposed that some aspects of localization behavior may well be mediated by the output of the population of these neurones [for example, see Middlebrooks (1984)], in which case, the neuronal point image, or extent of the nucleus which was activated by a single point stimulus in space, would vary significantly between these species. We were interested in varying the ‘‘point image’’ of the output layer in our model to see what effect this may have on the subsequent localisation accuracy of the model.

The preprocessing components of the model described in this report attempts to account for (i) the spectral smoothing of cochlear encoding, (ii) the frequency-dependent variation in auditory sensitivity, (iii) the encoding of acoustical information as spike trains, and (iv) the parallel streaming of interaural timing information and spectral amplitude information. Once a combination of parameters had been determined that produced human-like localization performance, two main experiments were conducted using the model. First, we explored the fidelity of the spectrum amplitude quantization of the input required to sustain human localization performance. Second, the impact of the size of the ‘‘point image’’ of the ANN output was explored by varying the extent of the ANN output layer over which the spatial interpolator took its input. A third control experiment was also carried out to ensure that the behavior of the ANN was not limited by the information encoding capacity of the network architecture. In this case the effects of varying the number of hidden layer neurones on localization performance was examined.

I. STRUCTURE OF THE LOCALIZATION MODEL

A. Model overview

The model consists of three parts. First, a broadband sound in free field space was simulated using white noise [Fig. 1(a)]. The noise was filtered by the filter function of the outer ear, or head-related transfer function (HRTF), adjusted using the frequency sensitivity function of the auditory system and then used as input to the Patterson–Holdsworth cochlear model (Slaney, 1994). The monaural and binaural localization cues (left and right monaural spectral cue and ITD cue) were extracted from the cochlear output [Fig. 1(b)]. An artificial neural network (ANN) was trained to classify these localization cues [Fig. 1(c)]. The model was developed using the Matlab scripting language (version 4) and the C programming language. In this model, the direction of a sound source with respect to the listener is given using the vertical, single-pole coordinate system with azimuth 0 degrees and elevation 0 degrees indicating the position directly ahead of the subject. Locations above the audio-visual horizon and to the right of the anterior of the midline are indicated by positive degrees elevation and azimuth, respectively (Carlile, 1996).

B. Stimulus generation and localization cue encoding

Due to the various level-dependent nonlinearities in auditory encoding, the capacity of the system to encode spectral shape is dependent upon the overall input level (Sachs and Young, 1979). This will also be modified by the frequency dependency of auditory sensitivity that reflects, in part, the acoustical transmission properties of the auditory periphery (the pinna, concha, ear canal, and middle ear). The minimum audible field (MAF) describes the minimum detectable pressure level determined at the position of the subject's head for a free-field, pure tone stimulus located on the median plane (ISO R.226; see also Glasberg and Moore, 1990). This variation of sensitivity should affect the audibility of different frequency components of a complex sound. In this context, the MAF will weight the spectral cues according to the human audiometric sensitivity so that some of the features of the HRTF will be more salient than the others.

Carlile and Pralong (1994) argued that the neural excitation pattern for a spectrally flat broadband noise directly in front of a subject (azimuth 0 degrees, elevation 0 degrees) could be estimated by passing the inverted MAF through a cochlear model. This study extended this method to estimate the neural excitation pattern for a sound at any location for which the HRTF had been determined. The MAF used in this model was taken from Glasberg and Moore (1990) and was assumed to correspond to the human sensitivity for a sound located at (0 degrees, 0 degrees) in our system. The MAF curve was extrapolated with a low-order spline to estimate the low- and high-frequency tails of the sensitivity function not covered by the original measurements. The neural excitation pattern for a location (az, el) was then estimated by passing a weighted spectrum $Ws_{(\mathrm{az,el})}$ for a broadband sound source at (az, el) to the cochlear model where $Ws_{(\mathrm{az,el})}$ was an approximation of the MAF at (az, el). The magnitude response of $Ws_{(\mathrm{az,el})}$ was estimated using the HRTFs measured at (az, el) and (0 degrees, 0 degrees) using Eq. (1):

$$|Ws_{(\mathrm{az,el})}(f)| = \left| \frac{k^{*}\mathrm{HRTE}_{(\mathrm{az,el})}(f)}{\mathrm{HRTF}_{(0°,0°)}(f)^{*}\mathrm{MAF}_{(0°,0°)}(f)} \right|. \qquad (1)$$

Equation (1) computes the magnitude response of $Ws_{(\mathrm{az,el})}$ by adjusting the MAF at location (0 degrees, 0 degrees) by the difference between the HRTF at (az, el) and (0 degrees, 0 degrees). The constant k adjusts the weighted spectrum to a reasonable level corresponding to a spectrum amplitude of 30 dB (i.e., $k=31.6$) (see also Carlile and Pralong, 1994). Since no phase information was included with the published MAF, it was assumed that the phase of $\mathrm{HRTF}_{(\mathrm{az,el})}$ and $Ws_{(\mathrm{az,el})}$ are identical [Eq. (2)]:

$$\arg(Ws_{(\mathrm{az,el})}(f)) = \arg(\mathrm{HRTF}_{(\mathrm{az,el})}(f)). \qquad (2)$$

A randomly generated white noise was filtered with $Ws_{(\mathrm{az,el})}$ and produced the input to the cochlear model needed to predict the neural excitation pattern for location (az, el). The duration of the noise stimulus was 100 ms.

The Patterson–Holdsworth cochlear model provided the simulation of the auditory transduction model using a gamma-tone filter bank and the Meddis hair cell model (Slaney, 1994). The output of the gamma-tone filter bank was connected to an array of Meddis hair cells. The compressive nonlinearity of the Meddis hair cell model limited the output of the cochlear model. The output of the hair cell gives the firing probability at each sampling interval. Each cochlear model contained 128 Meddis hair cells. The number of hair cells in the model is a trade-off between spectral resolution and computational efficiency and our choice was influenced by the model of Neti *et al.* (1992) who used the same number of channels to represent a spectral cue.

C. Extraction of ITD cues

Figure 1(b) shows the circuit for the extraction of the ITD and the left and right monaural spectral cues encoded in the output of the cochlear model. Figure 2 shows the circuit that extracts the ITD cue from the left and right cochlear outputs. This design was based on the silicon model of the time-coding pathway of the owl described in Lazzaro and Mead (1989) which operates on binary pulse/trains. The modified ITD circuit calculates the cross-correlation coefficients from the profile of the left and right inputs to the circuit.

The output of the hair cells in the left and the right cochlear outputs with the same center frequency were passed through a pair of time delay lines. Assuming that the ears are two points on a spherical head, the path difference d between the two ears is approximated by $d = r(\theta + \sin\theta)$ where r is the radius of the head and θ is the direction of the sound source measured from the median plane in radians (Woodworth, 1938). The maximum path length-occurs at $\theta = \pi/2$. From this equation the maximum path difference d is 0.26 m for $r = 0.1$ m and corresponds to a time delay of 780 μs. This defines the maximum time delay to be modelled for a signal to propagate from the start to the end of the delay line. The time delay elements updated their output by copying the output of the preceding time delay element to its own output at a rate of 40 kHz (i.e., $\Delta t = 25\,\mu\mathrm{s}$). Each time delay line contained a chain of 32 determined by the relationship between the maximum output rate of the cochlear model (40 kHz) and the need for the number of time delay elements to span the maximum ITD (i.e., $N = 40\,000 \times 780\,\mu\mathrm{s} \approx 32$). In addition, 32 delay elements would quantize the ITD to approximately 25 μs which is similar to the human just noticeable difference (JND) for an ITD of 10 μs (Yost, 1974). The output of all hair cells propagated through the time delay elements at the same rate. The cross correlation was calcu-

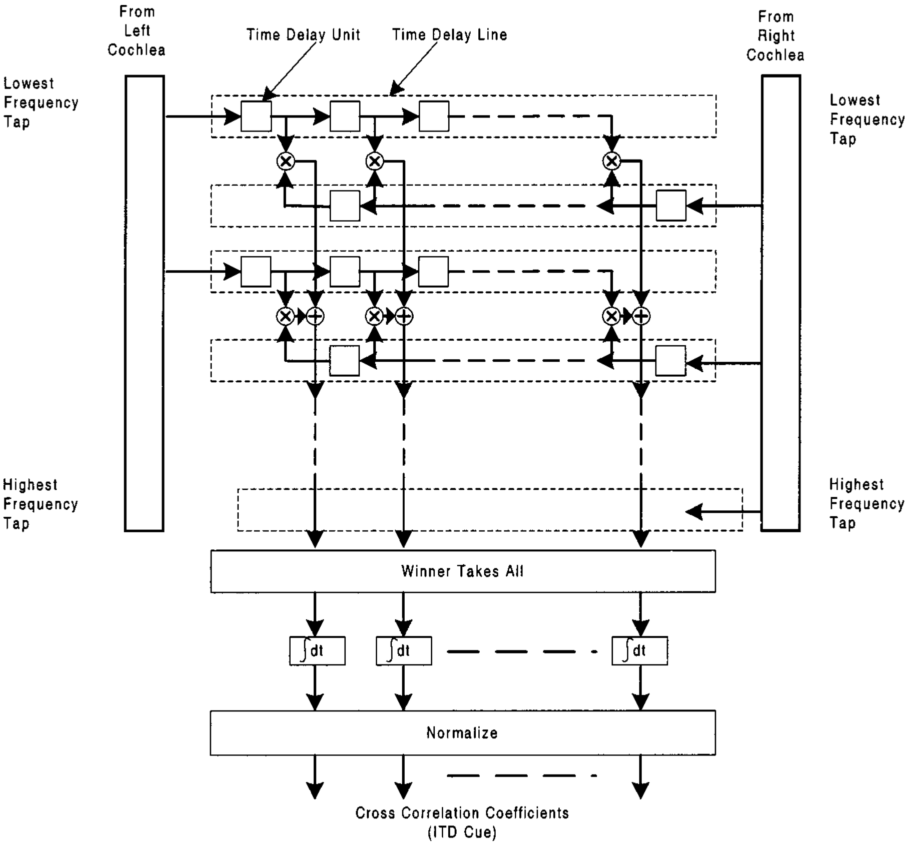

FIG. 2. The ITD cue extractor.

lated by multiplying the output of cochlear models passing through the two time delay lines [Eq. (3)], where cgm_l and cgm_r are the left and right cochlear outputs and N is the number of delay elements in a delay line:

$$\begin{aligned} xcorr_k(t) = & \sum_{f=100\ \mathrm{Hz}}^{1.4\ \mathrm{kHz}} cgm_l(f, t-k\Delta t) \\ & \times cgm_r(f, t-(N-k-1)\Delta t) \\ & \text{for } 0 \leqslant k < N. \end{aligned} \tag{3}$$

The ITD cue extractor summed the cross-correlation coefficients for all hair cells with center frequency below 1.4 kHz as described in Eq. (3). A ''winner takes all'' function converted the cross-correlation coefficients into a binary vector $\mathrm{ITD}(t)$, where the ith component of this vector, $\mathrm{ITD}_i(t)$, was computed using Eq. (4):

$$\mathrm{ITD}_i(t) = \begin{cases} 1 & \text{if } xcorr_i(t) \geqslant xcorr_j(t) \quad \text{for all } 0 \leqslant j < N, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Bit i in the binary vector $\mathrm{ITD}(t)$ was set to one if it corresponded to the maximum cross-correlation coefficient, otherwise it was set to zero. The binary versions of the instantaneous ITD cues were accumulated [Eq. (5)] to estimate the normalized ITD cue [Eq. (6)]. In most cases when the model was stimulated by a broadband stimulus, only one component in $\mathrm{ITD}_{\mathrm{norm}}$ was set to one while the other components were close to zero:

$$\mathrm{ITD}_i = \sum_{t=25\ \mathrm{ms}}^{100\ \mathrm{ms}} \mathrm{ITD}_i(t) \quad \text{for } 0 \leqslant i < N, \tag{5}$$

$$\mathrm{ITD}_{\mathrm{norm}} = \left(\frac{\mathrm{ITD}_0}{\mathrm{ITD}_{\max}}, \frac{\mathrm{ITD}_1}{\mathrm{ITD}_{\max}}, \ldots, \frac{\mathrm{ITD}_{N-1}}{\mathrm{ITD}_{\max}} \right)$$

$$\text{where } \mathrm{ITD}_{\max} = (\max \mathrm{ITD}_0, \ldots, \mathrm{ITD}_{N-1}). \tag{6}$$

Within a particular frequency channel, the ITD information corresponded to the phase relationships between the signals received from the left and right ears. As a result, only frequencies with a wavelength greater than the distance between the ears offer an unambiguous ITD.

D. Extraction of spectral cues

The monaural spectral cue for the model was the time average of the cochlear output from 25 to 100 ms. The first 25 ms of output was discarded to allow the cochlear output to reach a steady state. Note that the left and right monaural spectral cues were extracted simultaneously and presented to the ANN together. This arrangement allowed the ANN to derive binaural cues such as the interaural spectral difference cue or the interaural level difference cue using the two monaural spectral cues.

E. Artificial neural network classifier and feature vector coding

A two-layer artificial neural network (Rumelhart and McCelland, 1986; Lippmann, 1987) was trained to transform the localization cues to a two-dimensional output matrix that indicated the direction of the sound source. A feed-forward

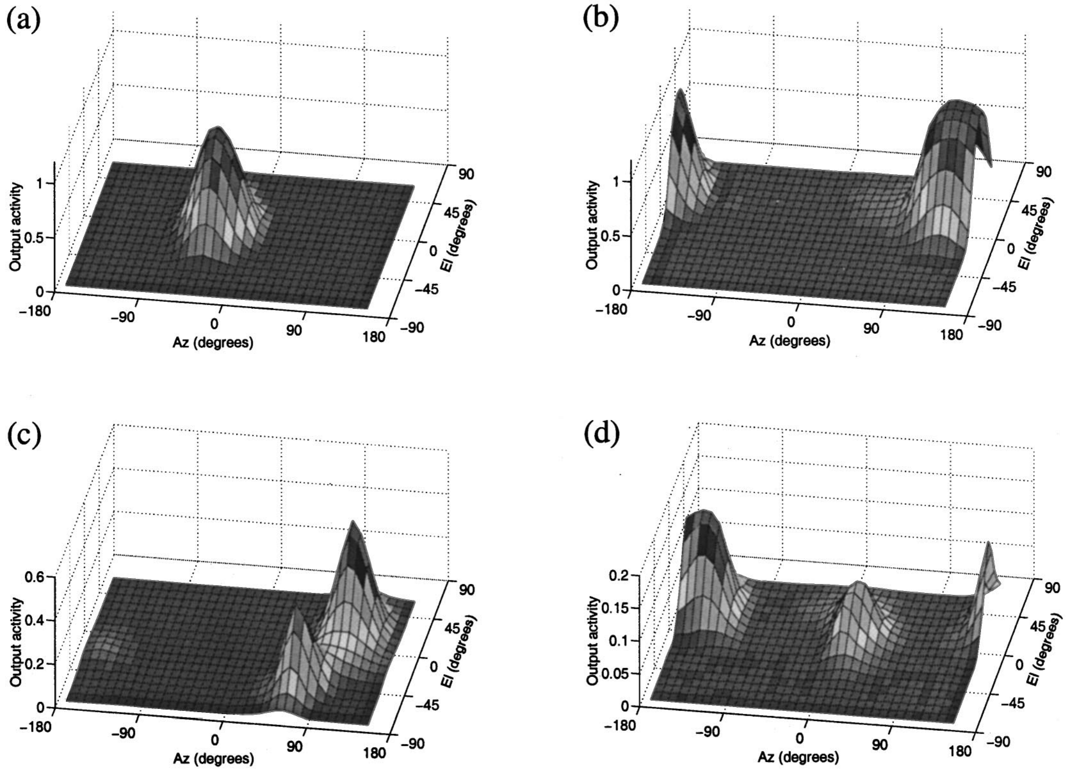

FIG. 3. Examples of ANN outputs. (a) A single clear peak for an estimated source located at (−40 degrees, 0 degrees). (b) A single clear peak for a source estimated to be at (140 degrees, 40 degrees). (c) Output of ANN which led to elevation confusion with source located at (80 degrees, −40 degrees). (d) Output of ANN which led to front–back confusion with source located at (40 degrees, 20 degrees). Data points were interpolated at 10 degrees rather than 1 degree as described in Sec. I E for clarity of this illustration.

back-propagation algorithm was used to train the network. A neural network simulator, NevProp (Goodman *et al.*, 1993), was used to implement the neural network.

The ANN contained 288 input neurones: 128 neurones to receive the left spectral cue, another 128 neurones to receive the right spectral cue, and 32 neurones to receive the normalized ITD cue. The model used in the experiment that was described in Sec. III A contained eight hidden layer neurones. The output layer contained 162 neurones which formed a 9×18 neurone matrix. Each neurone in the matrix was associated with a point (az,el), where $-180 \leqslant \mathrm{az} < 180$ degrees and $-80 \leqslant \mathrm{el} \leqslant 80$ degrees, with a 20-degrees step size. The neurones in the input and the hidden layer were fully interconnected, and likewise for the neurones in the hidden layer and the output layer. There were no assumptions about how the network would use the localization cues. Although the two poles (north and south) were not represented directly by any neurones, the localization model could still specify these directions by passing the ANN outputs to a space-map interpolator (described below).

Each training vector presented to the ANN consisted of two parts: the ANN input vector and the target output vector. The input vector consisted of the left and right spectral cues and the ITD cue for direction (az,el). The left and right spectral cues in each input vector were scaled by the normalization factor $1/m$, where m was the maximum value in each pair of left and right spectral cues. The target output vector specified the output activity of the output layer neurones during training. When the neurone output activity is plotted over its associated coordinates, the profile of the plot resembles a double Gaussian distribution. The coordinates with the highest output activity was taken as the source direction. Equation (7) was used to calculate the output activity for an output neurone d_{ij} located at position ($20i$ degrees, $20j$ degrees) where $-9 < i \leqslant 9$ and $-4 < j \leqslant 4$ for a sound source located at (az, el). The variable σ is the standard deviation of the double Gaussian distribution. A large value of σ causes more output neurones to respond to an input stimulus. The value of σ was set to 30 degrees in all model experiments unless otherwise stated:

$$d_{ij} = e^{-[(20i^\circ - \mathrm{az})/(\sigma/\cos(20j^\circ))^2 + ((20j^\circ - el)/\sigma)^2]}. \tag{7}$$

The single pole representation of space used in the study has the disadvantage that as elevation diverges from the greater circle indicating the equator, there is a decrease in the area of space corresponding to one degree azimuth. A scaling factor, $1/\cos(20j)$ (where $20j$ is the elevation associated with the output neurone) has been included in Eq. (7) to adjust the spread of the activity to compensate for the change in density of neurones under a fixed area at different elevations.

The ANN output was interpolated using the Matlab function griddata to give the final location estimate. The algorithm interpolates the ANN outputs from a 20-degree grid to a 1-degree grid. The most active output neurone was identified and the interpolation operation was applied to the local 3×3 output neurone matrix centered on the most active output neurone. Figure 3 illustrates four examples of the

ANN outputs. The plots show the pattern of activity of the output layer which reflects stimulation at four different points in space. Note that the maximum output activity shown in Fig. 3(a) and (b) is close to 1, whereas in Fig. 3(c) and (d) the maximum output activity is lower. In the latter two cases, the ANN had difficulty in estimating the source location, and generated output activities for two regions which indicates an increase of elevation [Fig. 3(c)] or front–back confusion [Fig. 3(d)] errors.

II. EXPERIMENT

A. Psychophysical test setup

Since the model was designed to deliver human level performance, it was necessary to assess the model performance against the free-field localization performance of the human listener who provided the HRTFs to the model. The free-field localization tests have been described in detail elsewhere (Carlile *et al.*, 1997) but are briefly outlined below. All localization tests that involved human listeners were carried out in the same anechoic chamber that was used for the HRTF recording (Carlile *et al.*, 1997). A 150-ms broadband stimulus was played through a loudspeaker mounted on a computer-controlled, semi-circular hoop (1-m radius) centered on the listener's head. The listener was asked to point his/her nose toward the perceived location of the sound source. An electromagnetic tracking device was mounted on the listener's head which measured the position of the head. This procedure was repeated for all 76 locations with four trials per location. The localization estimates were analyzed using the same procedures applied for the model's results for comparison.

B. Data analysis

The azimuth and elevation systematic errors, the spherical angular error of localization estimates, and the overall spherical correlation coefficient (SCC) were calculated for the model results to assess the performance of the model. The randomness in the stimuli supplied as input to the model would certainly cause some variations in the estimated source location, and therefore multiple estimates for each source location were collected. The difference between the source position and the mean direction of the distribution gives the systematic error in azimuth and elevation directions (referred to as E_{az} and E_{el}, respectively) (Leong and Carlile, 1997). For the set of estimates made for the same source location, the spherical angular error is defined as the average of the angle between the line joining the center of the head of listener to the mean direction and the line joining the head of the listener to each of the estimates.

The spherical correlation coefficient is a measure of the correspondence between the actual location of the target and the location indicated by either the human subjects (Carlile *et al.*, 1997) or the model. Consistent with other studies we have used this coefficient as a global metric of localization accuracy (for methods of calculation see Wightman and Kistler, 1989; Fisher *et al.*, 1993; Carlile *et al.*, 1997). Front–back confusion errors were processed separately during data analysis. An estimate was considered front–back confused if the source and the estimate were on different sides of the interaural axis and greater than 5 degree from the vertical plane through the interaural axis. In addition, an estimate was considered front–back confused only if it lay within ±20 degrees azimuth of the mirrored source location. This operation was employed to separate the large azimuth errors from front–back confusion errors. All data analysis was performed using the Matlab toolbox, SPAK (Leong and Carlile, 1997).

C. Vector generation

The HRTFs were used to simulate sound sources located at different positions in space. The HRTF library of a human subject was recorded using the procedures described in Pralong and Carlile (1996). A total of 724 pairs of HRTFs were recorded for equally spaced stimulus positions.

Localization cue vectors were generated by passing filtered noise stimuli through the cue coding and extraction components of the model as described in Sec. I. A range of input levels were prepared: 40, 50, 60, 65, 70, 80, and 90 dB. In addition, at each input level, ten samples were generated for each pair of HRTFs contained in the 724-point HRTF library. This provided a vector pool that contained a total of 50 680 localization cue vectors. Training and testing vector sets were formed by selecting localization cue vectors from this large pool that corresponded to the particular stimulus characteristics of interest. Humans are most likely trained to localize on a wide range of stimulus levels. This was simulated in the training of the model by using training cue vectors generated at different input levels. The training vector set was composed of one sample per point in the HRTF library for each input level. This gives a total of 5068 cue vectors in the training set. The testing vector set was composed of a subset of 76 positions that were used in the training, with ten samples of localization cue vectors generated at 65-dB input level. The performance of the model on this test set was compared with the localization results of human listeners using the same set of 76 stimulus locations. The human psychophysical experiments were performed at a fixed 65-dB sound pressure level (SPL) and used the same set of stimulus locations as that used in the generation of the testing vector set.

III. RESULTS

A number of parameters were available to control the performance of the model which was then compared to the performance of the single subject who provided the HRTFs used in the model as well as the performance data pooled from a population of 19 other subjects. Localization performance was measured in terms of the spherical correlation coefficient of the localization estimates against the actual target locations and the percentage of front–back confusion errors. We have included comparisons of both the single subject and the population response as it is likely that the performance of the single human subject also contains idiosyncrasies that reflect factors other than the sensory inputs. Therefore, a more robust comparison of the model performance might be against the population response where individual effects should have been cancelled out to some extent.

TABLE I. Localization performance of individual listener (Human), a group of 19 human listeners (Group), and the reported model with different configuration (E1 to E14). The test stimuli were generated at 65 dB SPL.

Result	Trial no.	Quantization levels	Cochlear channels	Hidden neurone	σ (degrees)	SCC (no MAF)	% fb (no. MAF)	SCC (MAF)	% fb (MAF)
Human	304	n/a	n/a	n/a	n/a	n/a	n/a	0.957	1.0%
Group	6909	n/a	n/a	n/a	n/a	n/a	n/a	0.983	3.2%
E1	780	10	128	8	30	0.958	8.1%	0.884	7.5%
E2	780	20	128	8	30	0.966	3.7%	0.944	3.3%
E3	780	40	128	8	30	0.962	5.9%	0.972	0.8%
E4	780	100	128	8	30	0.970	5.4%	0.975	1.5%
E5	780	20	64	8	30	0.934	7.6%	0.950	4.1%
E6	780	20	256	8	30	0.969	3.3%	0.974	2.7%
E7	780	20	128	6	30	0.883	7.2%	0.823	6.7%
E8	780	20	128	10	30	0.964	4.7%	0.956	3.7%
E9	780	20	128	8	5	0.274	33.7%	0.477	12.1%
E10	780	20	128	8	10	0.925	9.9%	0.886	4.6%
E11	780	20	128	8	20	0.957	4.1%	0.984	0.4%
E12	780	20	128	8	40	0.969	6.5%	0.968	1.6%
E13	780	20	128	8	60	0.955	3.1%	0.941	2.1%
E14	780	20	128	8	90	0.939	3.1%	0.948	0.0%

In Table I the summary results for all 28 experiments with the model are presented. Experiments E1–E4 examined the variation of the spectral cue amplitude resolution; E2, E5, and E6 examined the effects of varying the frequency resolution of the input to the ANN; E2, E7, and E8 examined variation in the number of neurones contained in the ANN hidden layer; and E2 and E9–E14 examined the variation in the size of the neural ''point image'' of the output. All experiments were conducted both with and without the MAF correction.

A. Broadband localization

The objective of the first series of experiments was to develop a model with the most human-like localization performance. In this series, we systematically varied the frequency resolution [64 (E5), 128 (E2), and 256 (E6) cochlear channels] and the size of the neural point image (5, 10, 20, 40, 60, and 90 degrees: E9–14). The SCC provides only a rough guide for comparisons of localization accuracy between the subject group and the model as this measure is sensitive to the number of samples. Therefore, as an initial screening of performance, we sought a combination of parameters that provide a SCC that was close to that of the individual listener and a front–back confusion rate that was closer to the group. This was provided by a model with 128 channels and a neural point-image of 30 degrees. For each model we also examined the azimuth and elevation systematic errors and the spherical angular error and these also demonstrated the best fit to the human data for our standard model. These measures have more explicit and detailed measures of localization performance, offering more information than the SCC.

The ''standard'' model result and the individual and group human results were analyzed and plotted as a function of the azimuth of the sound source. The plots include the azimuthal and elevational systematic errors (Figs. 4 and 5), the spherical angular error (Fig. 6) of the localization estimate distributions at ±40-, ±20-, and 0-degrees elevation, as well as the distribution of the front–back confusions (Fig. 7). On average the error magnitudes were very similar among the three sets of results.

The azimuthal systematic errors for the individual listener showed more abrupt changes than the data obtained from the group. For instance, Fig. 4 shows that at elevation 40 degrees, the azimuthal systematic error for the individual listener (circles) changed from 10 degrees at azimuth −150

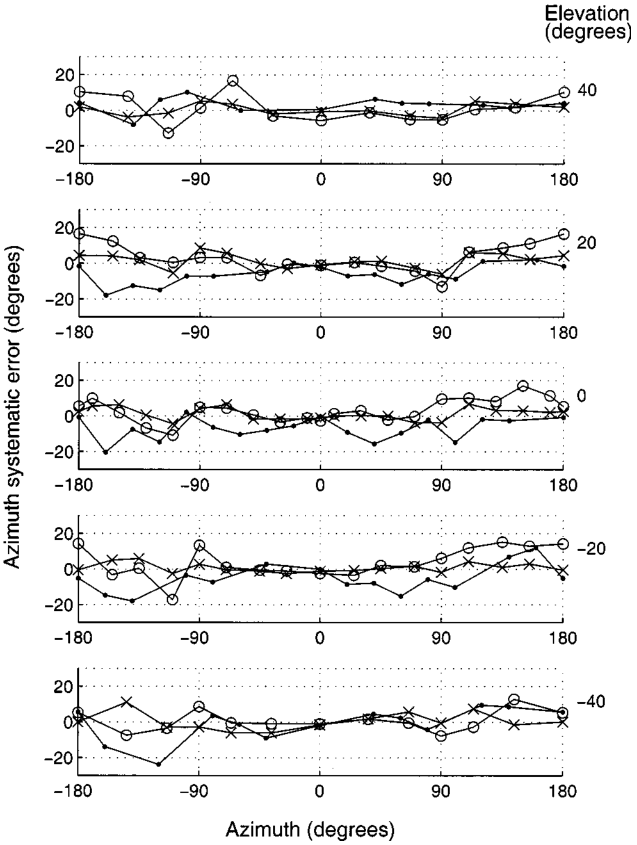

FIG. 4. E_{az} of the estimate distribution of the model (dot), the human listener whose HRTFs are used by the model (circle), and the group of 19 listeners (cross).

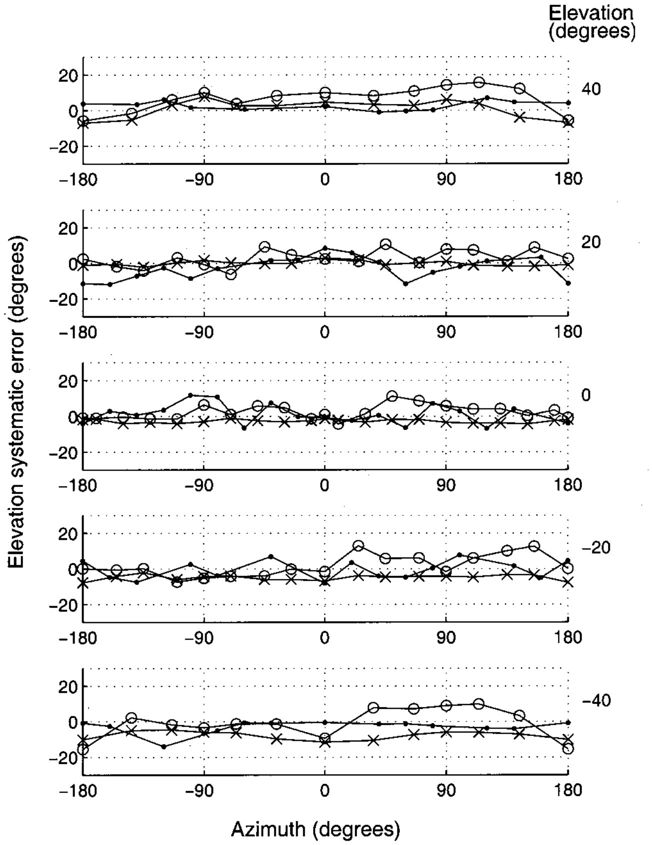

FIG. 5. E_{el} of the estimate distribution of the model (dot), the human listener whose HRTFs are used by the model (circle), and the group of 19 listeners (cross).

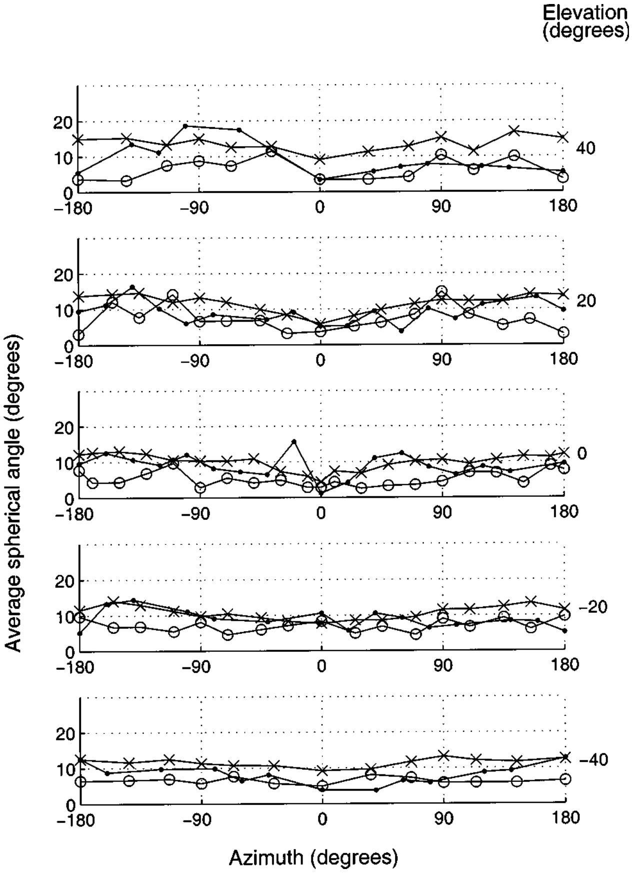

FIG. 6. Spherical angular error of estimate distribution of the model (dot), the human listener whose HRTFs are used by the model (circle), and the group of 19 listeners (cross).

degrees to −15 degrees at azimuth −110 degrees, then back to 20 degrees at −70 degrees.

The elevational systematic error is shown in Fig. 5 and indicates that the magnitudes for the elevational systematic error were similar for the three sets of data. At higher elevations, for example at 40 degrees, the magnitude of the elevational systematic error of the model (dots) was generally lower than those of the listener (circles) and the group results (crosses). The individual listener gave greater elevational systematic error at ±40 degrees elevations than at ±20 degrees and 0 degrees elevations, and often had positive elevational systematic errors when the source was on the right hemisphere. This feature was not found on the left hemisphere, and also not obvious in the model or the group result. This may represent a bias related to the response measure rather than the perception of the location.

Figure 6 shows the magnitude of the average spherical angular error of the estimates for the model, single subject, and group. The spherical angular error between the model (dots) and the individual listener results (circles) are similar at 20, −20, and −40 degrees elevation. The spherical angular error of the group (crosses) is generally greater than that of the model and the individual listener at all elevations. The average spherical angular error of the model remained within the limits defined by the individual listener and group results at most test locations.

The magnitude of systematic errors shown in Figs. 4 and 5 indicate that the individual listener exhibited performance that was close to the group results, and that the distribution of estimates were usually within the range of the group results. The model gave similar systematic errors for most locations to that shown for the single and group results.

The distribution of front–back confusion errors in the three sets of results with data collapsed across elevation are illustrated in Fig. 7. In the model result [Fig. 7(a)], the front–back confusion errors occurred more often with source locations on the left hemisphere. However, for the results of the group [Fig. 7(c)], the front–back confusion errors were more equally distributed for the right and left hemispheres.

B. Effects of frequency resolution

Increasing the number of cochlear channels from 64 (E5) to 128 (E2) resulted in only minor differences in the performance of the model, but that an increase to 256 (E6) channels resulted in a substantial increase in accuracy. In any event, the model with 128 cochlear channel provided the best match to the human performance.

C. Effects of amplitude resolution

To explore the role played by the fine structure of the spectral cues in localization performance, the spectral amplitude components of the monaural spectral cues were subjected to a linear quantization operation. The magnitude of the quantization step applied to the monaural spectral cues was varied and four models were trained using spectral cues

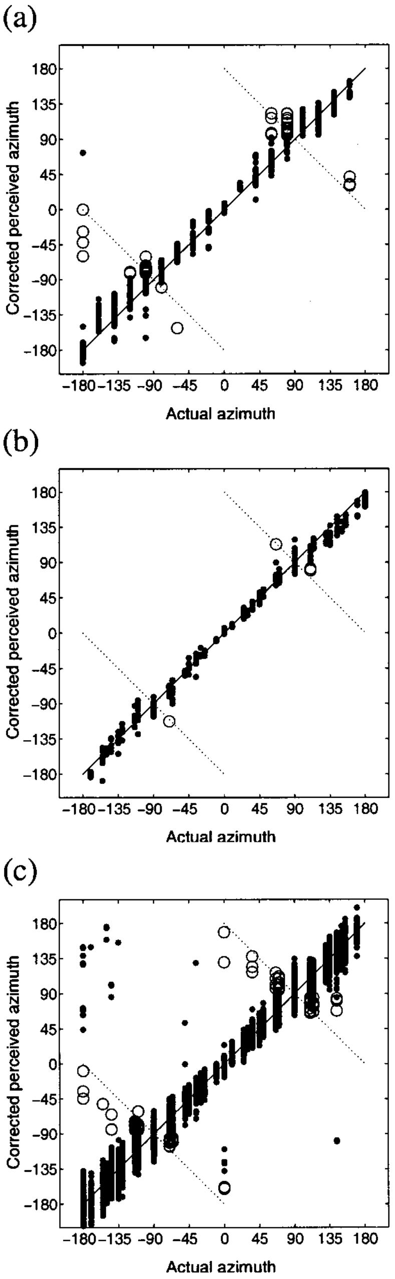

FIG. 7. Front–back confusions of the (a) model, (b) human listener, and (c) the group of 19 listeners. Front–back confused data are plotted using empty circles and non-front–back confused data are plotted using dots.

that were quantized to 10, 20, 40, and 100 levels over the total dynamic range of the normalized spectral cues. For instance, when the quantization parameter was set to 40 levels, a normalized monaural spectral cue was approximated by 40 equally spaced discrete values between 0 to 1 (inclusive). Each analog value of spectrum amplitude was rounded to the closest quantization level. The azimuth and elevation systematic errors (Figs. 8 and 9), spherical angular error (Fig. 10), and the front–back confusion errors (Fig. 11) have been plotted for 10, 20, and 40 quantization levels. E1–E4 in Table I correspond to the models trained with spectral cues quantized to 10, 20, 40, and 100 quantization levels, respectively.

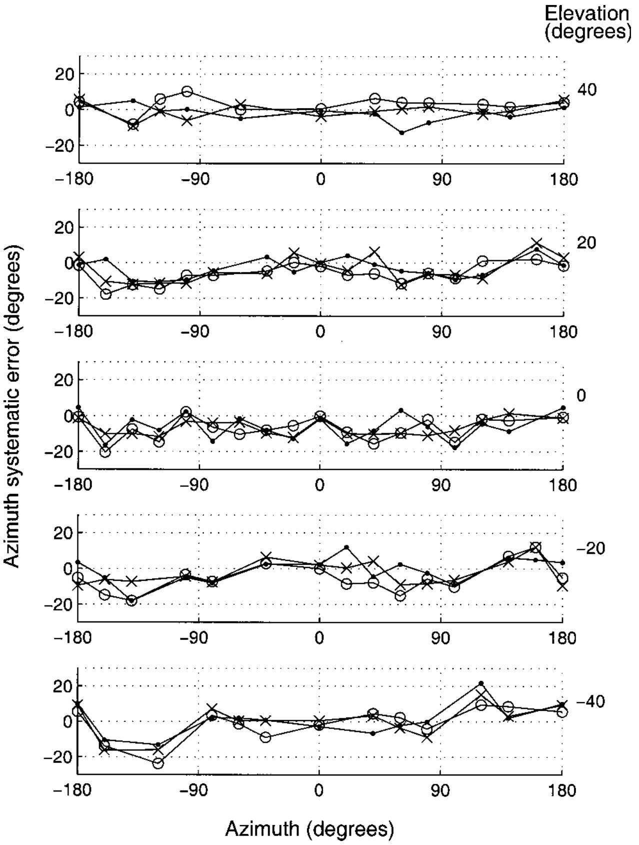

FIG. 8. E_{az} for models trained with spectral cues quantized to 10 (dot), 20 (circle), and 40 (cross) levels.

The pattern and magnitude of the systematic errors demonstrated by the different models showed only marginal changes with variation in the number of quantization levels (azimuth: Fig. 8 and elevation: Fig. 9). The random errors of localization indicated by the extent of the distribution of the estimates showed clear changes when the number of quantization levels were changed (Fig. 10). At all elevations, the model that used spectral cues with higher amplitude resolution (greater number of quantization levels) gave smaller average spherical angles for the distributions. For the directions close to (0 degrees, 0 degrees), however, all three models had similar spherical angular error. In addition, the extent of the front–back (FB) confusions were also negatively correlated to the number of quantization levels (10 to 40 levels). From Fig. 11, the model with 10 quantization levels produced the most front–back confusion errors among the three models; the model with 40 quantization levels gave almost no front–back confused estimates.

The levels of performance with 20 quantized levels was closest to that exhibited by the human subjects (Figs. 4, 5, and 6). Increasing the number of quantized levels beyond 40 produced only a slightly higher SCC and more front–back confused estimates.

D. Effects of number of hidden layer neurone

The number of hidden layer neurones will affect the information storage capacity of the ANN (Baum and Haussler, 1989). As a control experiment, it was important in this study to demonstrate that the ANN architecture we were employing was not a limiting factor in the performance of the model. In experiments E7, E2, and E8, the number of hidden

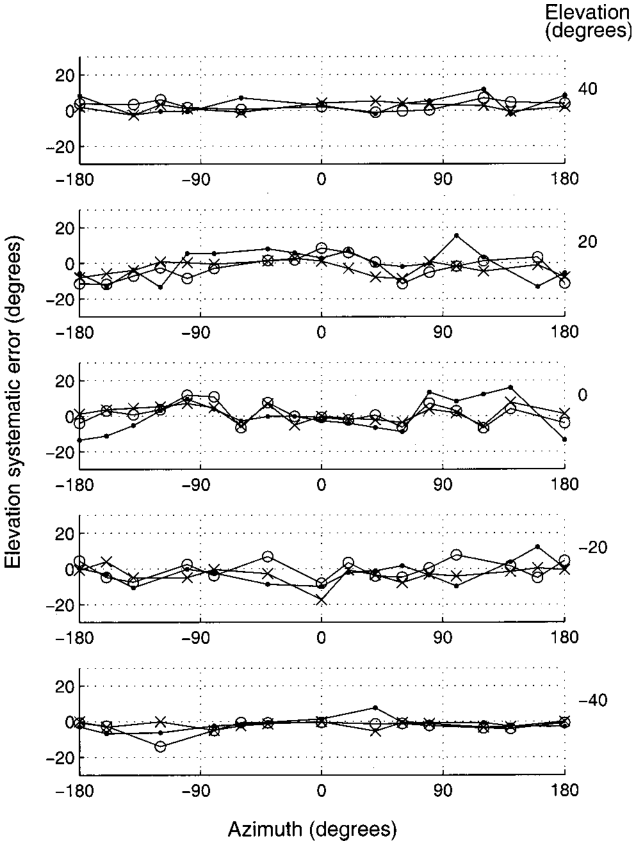

FIG. 9. E_{el} for models trained with spectral cues quantized to 10 (dot), 20 (circle), and 40 (cross) levels.

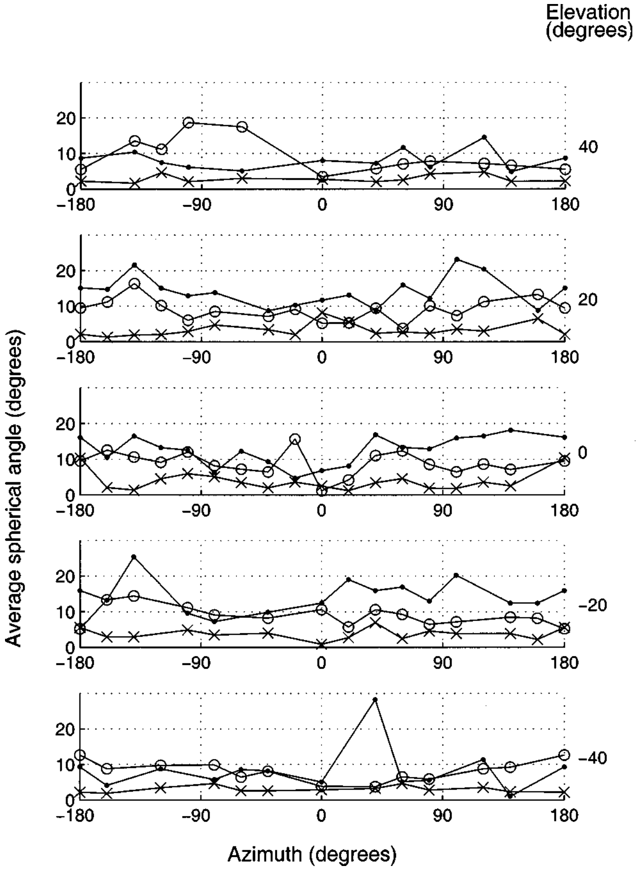

FIG. 10. Spherical angular error for models trained with spectral cues quantized to 10 (dot), 20 (circle), and 40 (cross) levels.

layer neurones was six, eight, and ten, respectively. With six hidden layer neurones, there was a substantial decrease in the SCC and a twofold increase in the number of FB confusions when compared to the model with eight hidden layer neurones. In contrast, an increase to ten neurones had only a marginal effect on the SCC and a similar number of FB confusions. We concluded from this experiment that the performance of our standard model was not bound by capacity of our ANN with eight hidden layer neurones required to carry out an effective input–output mapping.

IV. DISCUSSION

A. Broadband localization

The localization performance of the model was compared with the individual listener from whom the HRTFs were obtained and with a group of 19 human listeners (Carlile *et al.*, 1997). Although there is a good match between the three sets of data, there are some characteristic differences. Some of these differences can probably be attributed to the differences in the sample sizes for each group. This is particularly evident with the sharp variations in the azimuthal systematic errors (Fig. 4) for the individual human data compared to the group data and to a lesser extent the model.

A second source of difference is likely to originate from the response measures. In testing human localization performance, the listener was asked to point his or her nose in the direction of the source (Carlile *et al.*, 1997). Although turning to face towards a sound source is a highly ecological process, there are at least two types of errors that may be present in the estimates. First, for sound locations that require very small or large movements of the head to the mechanical limits of motion, there is a tendency for the listener to point towards the source using a combined movement of the head and the eyes, rather than only the movement of the head. As the method of detecting the perceived location of the sound source involves tracking the position of the head (rather than the eyes), the ‘‘eye capture’’ of the perceived target location may produce systematic errors in estimating the perceived location. This type of systematic error is classified as a motor error (for a full discussion see Carlile *et al.*, 1997). In addition to motor error, there are also spatially dependent variations in the localization performance found in experiments that have minimized the presence of motor errors (Gilkey *et al.*, 1995; Carlile *et al.*, 1997) which can be attributed to sensory error. Since the model used the HRTFs of the individual listener, one would expect similar features to be found in their results if the behavior of the model was constrained similarly to that of the human even though the magnitudes of the model systematic errors were close to the individual listener and the group results. The systematic errors (Figs. 4 and 5) made by the model were more symmetrical about the median plane than that of the individual listener. The asymmetry in the single listener’s errors might have resulted from motor errors or other response biases. We have not attempted to include any of these latter kinds of errors in our model.

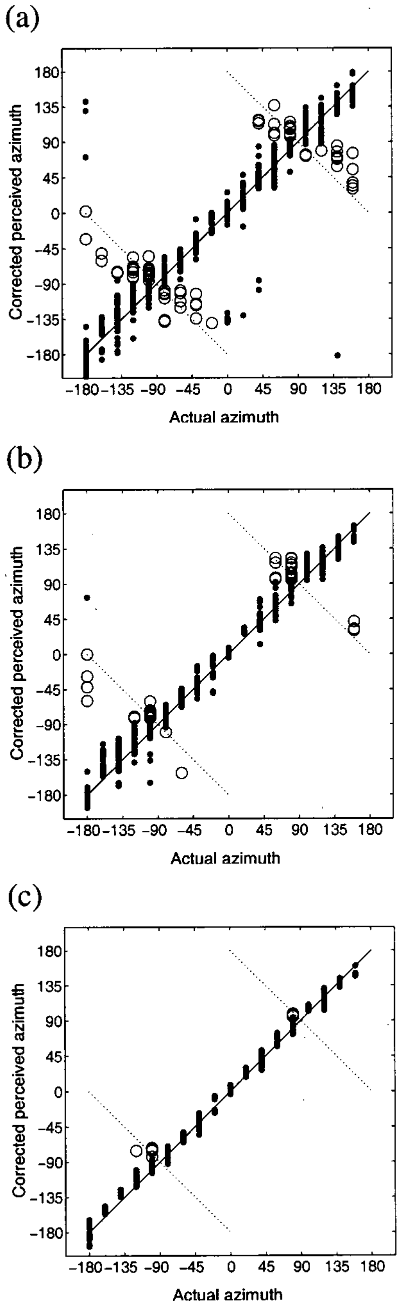

FIG. 11. Front–back confusions for models trained with spectral cues quantized to (a) 10, (b) 20, and (c) 40 levels.

For instance, both the systematic error and the size of the error distributions were observed to increase with the azimuth of the sound source. In localization experiments that involved a human listener, when the sound source was behind the listener, the listener was unable to swivel his/her head towards the back without body movement to point their nose at the perceived sound location. The judgment of the azimuthal angle traveled may well be less accurate than relatively smaller azimuth movements accomplished by the head alone. This is consistent with observations that for the human listeners, the azimuthal localization error was smallest when the source was on the median plane and close to the audiovisual horizon. The slightly greater accuracy demonstrated by the model for more lateral and posterior location may be due to the absence of motor error in the model's result.

The close similarities between the average spherical angular error and systematic error close to the median plane in the results collected from the model and the human listeners also suggests that localization cue features used by the model were similar to those used by human listeners. This in turn provides support for the notion that the model parameters were set to psychophysically plausible values.

B. Effect of amplitude quantization on localization performance

There was no significant change in the pattern of the systematic errors with the variations in the number of quantization levels applied to the monaural spectral cues. Rather, the more significant changes were in the magnitude of the distribution of the estimates. In general, the spherical angular error of the distribution of estimates increased as the number of quantization levels was decreased. This indicates that the detail in the monaural localization cues can provide information important for accurate localization performance. It was interesting that even with a quantization level of only ten steps, the extent of the systematic errors was similar to the other models, and the distribution of estimates was only slightly greater than models with higher quantization. However, with ten-step quantization, there was a significant increase in front–back confusions.

This strongly suggests that the detail in the spectral cues could contribute to the processes involved in resolving front–back confusions. Of interest, the model with 100 quantization levels showed slightly more front–back confusions than the model with 40 quantization levels although the SCCs at 40 and 100 were the same. This may indicate that the fine detail in the spectral cues caused by the randomness of the white noise had an undue influence on the development of the network during training. As a result the model became more sensitive to such random detail which resulted in increased front–back confusions.

This result suggests that the auditory system may not need to exploit the full detail of the spectral cues available to perform accurate localization. Alternatively, the auditory system may be unable to encode or resolve high resolution changes in amplitude that are characterized by the measured acoustical HRTF. From the experiments, the model that used spectral cues with 20 discrete levels produced a performance most similar to that of the human, which implies that the auditory system might only require about 4 bits of resolution on the spectral amplitude cues to perform reliable localization.

C. Effect of MAF correction on model performance

An initial design decision in the model front end was to include the frequency sensitivity function of the auditory system in the form of the minimum audible field (MAF) function in line with earlier excitation pattern estimates of Carlile and Pralong (1994). The underlying rationale was that the MAF weighting should affect the saliency of various spectral components of the broadband sound. The cochlear model itself does not have a frequency dependence in threshold sensitivity; however, the use of the MAF, determined

TABLE II. Model architecture summary.

	Neti *et al.* (1992)	Janko *et al.* (1995)	Chau and Duda (1995)	This work
Spectral analysis	n/a	FFT	FFT and gammatone filter bank	Patterson–Holdsworth cochlear model
Spectral cues	Yes	Yes	Yes	Yes
Interaural time difference cue	No	Yes	Yes	Yes
Interaural level difference cue	No	Yes	Yes	No
ANN	128 input 8 hidden 187 output	1 to 23 input 50 hidden 30 output	No	288 input 8 hidden 162 output
Azimuth and elevation range	$-75\leq az \leq 75$ degrees $-30\leq el \leq 90$ degrees	$-180< az \leq 180$ degrees $-54\leq el \leq 54$ degrees	Not stated	$-180< az \leq 180$ degrees $-80\leq el \leq 80$ degrees
Distortion on localization cues	No	Amplitude and time jitter	No	Quantization and MAF

psychophysically, lumps together all of the threshold sensitivity components of the auditory system. To test whether this weighting was relevant at the simulated stimulus levels used in this model we repeated the experiment using our standard model in the absence of MAF filtering. In the absence of MAF weighting, the SCC in experiment E2 increased from 0.944 (with MAF) to 0.966 (without MAF) but the front–back confusion weights were largely unchanged (3.3% with MAF to 3.7% without MAF). This suggests that, at the simulated stimulus levels used in training and testing the current model, the MAF has only a minor effect on the quality of the information available to the ANN. It is a matter for future investigation to explore the impact of the MAF at higher and lower input levels where the nonlinearity of cochlear encoding would more likely result in distortions of the encoded spectra.

D. Comparison with previously reported models

Several previously reported models have been reviewed (Searle *et al.*, 1976; Lazzaro and Mead, 1989; Middlebrooks, 1992; Neti *et al.*, 1992; Horiuchi, 1994; Chau and Duda, 1995; Lim and Duda, 1995; Janko *et al.*, 1995). Neti *et al.* (1992), Chau and Duda (1995), and Janko *et al.* (1995) employ models with a similar architecture to the model reported in this paper. Table II summarizes the architecture of these model and the model reported in this paper. Apart from the models by Lazarro *et al.* and Horiuchi which extracted only the ITD cue, the other models reviewed together with this work used separate processing streams to extract the localization cues (e.g., spectral cues and ITD cue).

All the models also used the HRTF to filter the acoustic cues into neural excitation patterns. Neti *et al.* (1992) used the HRTF measured in anaesthetized cats, and assumed that the HRTF and the neural excitation pattern had the same profile. Chau and Duda (1995) estimated the neural excitation pattern by filtering a broadband noise with the HRTF and passing the result through a cochlear model. However, neither model used the MAF to adjust the amplitude of the inputs to the model. All of the previously reported models used the full precision of the estimated neural excitation pattern, although Lim and Duda (1995) noted that the spectral fine structures were not necessary for accurate localization in their model. In contrast, we have explicitly examined the role of amplitude features using a quantization parameter which determines the number of equally spaced quantized levels available to code the monaural spectral cues.

The model reported by Neti *et al.* (1992) used an ANN for the transformation of auditory cues into a two-dimensional output map to indicate the position of the sound source. An important point of departure with this study was that Neti *et al.* employed HRTFs measured from cats. The ear of the cat is a quite different in acoustical structure to that of the human and is well modelled as a truncated conical horn (Calford *et al.*, 1984) whereas the human outer ear demonstrates more complex acoustic behavior (Shaw, 1974; Carlile, 1996, Chap. 2). As a consequence, the nature of the spectral cues to localization are likely to be quite different between these species. For instance, the HRTF of the cat displays a number of middle frequency notches whose center frequencies vary systematically with location (Rice *et al.*, 1992) where similar systematic changes in notches are not evident in the human HRTF (Carlile and Pralong, 1994). Not withstanding these differences in what is being modelled, the model itself provides important insights in the approach and we have adopted much of the architecture for the model reported here. For instance, all neurones in successive layers were fully connected, and the ANNs were trained to generate output activities which resemble a double Gaussian distribution when mapped to the two-dimensional output map. Sixty-four spectral channels were used as monaural inputs in Neti's model although we settled on 128 cochlear channels after exploring this parameter experimentally. Additionally we have included an ITD analysis stream which was not included by Neti *et al.* Their model with ten hidden layer neurones which was trained using both left and right monaural spectral cues gave the best results. This is consistent with the findings reported here.

The model reported by Janko *et al.* (1995) provided an amplitude and time jitter operator which is not found in other models. After filtering a sound with the HRTF, each point of the left and right channels were multiplied by a normally distributed amplitude jitter factor and subjected to a normally distributed time delay. The standard deviation of the amplitude jitter was 0.25 and the standard deviation of the time jitter was 20 μs. Application of quantization is similar to the addition of noise to the signal. When a neural excitation pattern is rounded to one of the closest ten equally spaced discrete amplitudes, the equivalent ‘‘amplitude jitter’’ is 5% of the maximum value, compared with 25% (0.25 standard deviation) used in Janko’s model. The model by Janko *et al.* also used an ANN to classify the localization cues and provide location estimates. In contrast, the ANN used in Janko’s model contained 50 hidden layer neurones although no output interpolator was used. The estimated azimuth was indicated by 1 of 24 output neurones and the estimated elevation by 1 of 6 output neurones.

E. General discussion of the model

This model has attempted to simulate the processes of localization cue encoding and extraction using a physiologically and psychophysically plausible preprocessor and a two-layer neural network. This model exhibited very similar localization performance to human listeners when the model parameters were set to a particular range of values.

When constrained using physiologically realistic parameters, the model described here achieved localization performance close to that of human listeners. Increasing the amplitude resolution of the monaural spectral cues beyond 20 quantized steps resulted in only small improvements in the model’s performance. It was also found that the model’s performance was positively correlated with the number of cochlear channels for a given number of quantization levels. The model with 256 cochlear channels and using a MAF correction showed a greater SCC than that found for the human listener. Together, these results indicate that the model’s localization performance is at human levels using monaural spectral cues with relatively low amplitude and frequency resolution when compared to the level of resolution that is commonly employed in recording the HRTFs. The results indicate that the model can operate on monaural spectral cues with lower amplitude resolution and still deliver similar performance if the cues have sufficient frequency resolution. The important implication here is that, even with considerable ‘‘degradation’’ of the input by the preprocessor, the fidelity or quality of the localization cue information is still very high.

In the reported model, there was no assumption on how the monaural cues (left and right spectral cues) and binaural cues were combined to estimate the source direction. The model was free to select a combination that minimized an error measure. The model was free to determine the limit of localization performance accuracy when there were no restrictions on the usage of localisation cues. Although we expect strong evolutionary pressures to maximize the use of the available localization cues, it is likely that other biological constraints ensure that the central nervous system is not as free in its processing. An ANN that uses a topology which reflects the structure of the central nervous system and incorporates other ‘‘realism constraints’’ (Zipser, 1992) into the design may give further insights into the usage and effect of the auditory cues.

ACKNOWLEDGMENTS

This research project was supported by the University of Sydney Research Grants Scheme, the Australian Research Council, and the National Health and Medical Research Council (Australia).

Baum, E. B., and Haussler, D. (**1989**). ‘‘What net size gives valid generalization?’’ Neural Comput. **1**, 151–160.

Calford, M. B., and Pettigrew, J. D. (**1984**). ‘‘Frequency dependence of directional amplification at the cat’s pinna,’’ Hear. Res. **14**, 13–19.

Carlile, S., ed. (**1996**). *Virtual Auditory Space: Generation and Applications* (Landes, Austin).

Carlile, S., and Pralong, D. (**1994**). ‘‘The location-dependent nature of perceptually salient features of the human head-related transfer function,’’ J. Acoust. Soc. Am. **95**, 3445–3459.

Carlile, S., Leong, P., and Hyams, S. (**1997**). ‘‘The nature and distribution of errors in the localization of sounds by humans,’’ Hear. Res. **114**, 179–196.

Chau, W., and Duda, R. O. (**1995**). ‘‘Combined monaural and binaural localization of sound sources,’’ in *29th Asilomar Conference on Signals, Systems, and Computers* (Asilomar, CA).

Fisher, N. I., Lewis, T., and Embleton, B. J. J. (**1993**). *Statistical Analysis of Spherical Data* (Cambridge UP, Cambridge).

Gilkey, R. H., Good, M. D., Ericson, M. A., Brinkman, J., and Stewart, J. M. (**1995**). ‘‘A pointing technique for rapidly collecting localisation responses in auditory research,’’ Behav. Res. Methods Instrum. Comput. **27**(1), 1–11.

Glasberg, B. R., and Moore, B. C. (**1990**). ‘‘Derivation of auditory filter shapes from notched-noise data,’’ Hear. Res. **47**(1–2), 103–138.

Goodman, P., Rosen, D., and Plummer, A. (**1993**). University of Nevada Center of Biomedical Modelling Research, Washoe Medical Center, H-166, 77 Pringle Way, Reno, NV 89520. Goodman@unr.edu

Horiuchi, T. (**1994**). ‘‘An auditory localization and co-ordinate transform chip,’’ Neural Information Processing 7, Abstract of Papers, p. 43. http://www.klab.caltech.edu/~timmer/respapers/respaper.html.

Janko, J., Anderson, T., and Gilkey, R. (**1995**). ‘‘Using Neural Networks to Evaluate the Viability of Monaural and Interaural Cues for Sound Localization,’’ in *Binaural and Spatial Hearing in Real and Virtual Environments*, edited by R. Gilkey and T. Anderson (Erlbaum, Mahwah, NJ), Chap. 26, pp. 557–570.

King, A. J., and Carlile, S. (**1994**). ‘‘Neural coding for auditory space,’’ in *The Cognitive Neurosciences* (MIT, Boston).

Lazzaro, J., and Mead, C. A. (**1989**). ‘‘A silicon model of auditory localization,’’ Neural Comput. **1**, 47–57.

Leong, P., and Carlile, S. (**1997**). ‘‘Methods for spherical data analysis and visualization,’’ J. Neurosci. Methods **80**, 191–200.

Lim, C., and Duda, R. O. (**1994**). ‘‘Estimating the azimuth and elevation of a sound source from the output of a cochlear model,’’ in *Proceedings of the 28th Asilomar Conference on Signals, Systems and Computers* (Asilomar, CA). Available at http://www-engr.sjsu.edu/~duda/Duda.Research.Refs.html

Lippmann, R. P. (**1987**). ‘‘An introduction to computing with neural nets,’’ IEEE Trans. Accoust. Speech Signal Process. ASSP **2**, 4–22.

Middlebrooks, J. C., and Carlile, S. (**1984**), ‘‘A neural code for auditory space in the cat’s superior colliculus,’’ J. Neurosci. **4**, 2621–2634.

Middlebrooks, J. C. (**1992**). ‘‘Narrow-band sound localization related to external ear acoustics,’’ J. Acoust. Soc. Am. **92**, 2607–2624.

Middlebrooks, J. C., and Green, D. M. (**1991**). ‘‘Sound localization by human listeners,’’ Annu. Rev. Phys. Chem. **42**, 135–159.

Neti, C., Young, E. D., and Schneider, M. H. (**1992**). ‘‘Neural network models of sound localization based on directional filtering by the pinna,’’ J. Acoust. Soc. Am. **92**, 3140–3156.

Oldfield, S. R., and Parker, S. P. A. (**1986**). ‘‘Acuity of sound localization: a topography of auditory space. III Monaural hearing conditions,’’ Perception **15**, 67–81.

Pralong, D., and Carlile, S. (**1996**). ‘‘The role of individualized headphone calibration for the generation of high fidelity virtual auditory space,’’ J. Acoust. Soc. Am. **100**, 3785–3793.

Rice, J. J., May, B., Spiron, G. A., and Young, E. D. (**1992**), ‘‘Pinna-based spectral cues for sound localization in cat,’’ Hearing Res. **58**, 132–152.

Rumelhart, D. E., and McCelland, J. L. (**1986**). *Parallel Distributed Processing: Exploration in the Microstructure of Cognition I* (MIT, Boston).

Sachs, M. B., and Young, E. D. (**1979**). ''Encoding of steady-state vowels in the auditory nerve: representation in terms of discharge rate,'' J. Acoust. Soc. Am. **66**, 470–479.

Searle, C. L., Braida, L. D., Davis, M. F., and Colburn, H. S. (**1976**). ''Model for auditory localization,'' J. Acoust. Soc. Am. **60**, 1164–1175.

Shaw, E. A. G (**1974**), ''The External Ear,'' in *Handbook of Sensory Physiology*, edited by W. D. Keidel and W. D. Neffs, Vol. VIII, Auditory System (Springer-Verlag, New York), pp. 455–490.

Shaney, M. (**1994**). *Auditory Toolbox: A MATLAB Toolbox for Auditory Modelling Work*, Apple Technical Report #45, Apple Computer, Inc., Advanced Technology Group.

Wightman, F. L., and Kistler, D. J. (**1989**). ''Headphone simulation of free-field listening. II: Psychophysical validation,'' J. Acoust. Soc. Am. **85**, 868–878.

Woodworth, R. S. (**1938**). *Experimental Psychology* (Holt, New York).

Yost, W. A. (**1974**). ''Discrimination of interaural phase differences,'' J. Acoust. Soc. Am. **55**, 1299–1303.

Zipser, D. (**1992**). ''Identification models of the nervous system,'' Neuroscience (NY) **47**, 853–862.