The low dimensional world of echolocating bats

Abstract

This template helps you to create a properly formatted LaTeX manuscript.

1. introduction

Echolocating bats are long-lived, mobile animals. They use the landscape intensively for foraging, roosting, and commuting. They have excellent spatial memory (Barchi et al., 2013; Von Helversen & Winter, 2005) and can rely on sonar to navigate to notable locations like foraging grounds and drinking places (see Vanderelst et al., 2016; Vanderelst & Peremans, 2017, for references). Bats deprived of sight have been found to successfully return to their roost when displaced by up to 40 miles (Stones & Branick, 1969; Williams et al., 1966; Davis & Barbour, 1970; Mueller & Emlen, 1957). Furthermore, bats have been reported to use sonar-based recognition of landmarks, both under experimental (Jensen et al., 2005; Yu et al., 2019), and natural (Verboom et al., 1999) conditions.

Research on animal navigation has revealed a range of potential strategies that can be used to reach a distant goal (Reviewed by Franz & Mallot, 2000). Irrespective of the strategy an animal uses, a mechanism for recognizing previously visited places is required (Vanderelst et al., 2016; Vanderelst & Peremans, 2017). This implies that echolocating bats can identify previously visited sites (or landmarks defining those places) using sonar. Currently, it is still an open question of how bats use sonar to recognize locations in the environment.

One mechanism for place recognition would be for bats to parse the stream of echoes returning from the environment into localized and identified objects (Lee et al., 2017; Barchi et al., 2013; Moss & Surlykke, 2001; Schnitzler et al., 2003; Simmons, 2012; Ulanovsky & Moss, 2008; Clare & Holderied, 2015; Surlykke et al., 2016; Geipel et al., 2013). However, as we argued before, it is highly questionable whether bats can interpret echoes in terms of a 3D environment model or acoustic image (e.g., Vanderelst, 2015; Vanderelst et al., 2016; Steckel & Peremans, 2013). Sonar projects the three-dimensional structure of the environment on two one-dimensional waveforms arriving at the ears. While the time of arrival of echoes encodes object distance, azimuth and elevation directions are not directly available from the acoustic signal. In principle, azimuth and elevation can be estimated using spectral cues imposed by the head-related transfer function (e.g., Reijniers et al., 2010; Aytekin et al., 2004). However, complex environments, and even single plants (Yovel et al., 2008; Müller & Kuc, 2000), return many overlapping and interfering echoes, making reconstructing the direction of origin of echoes much harder, and perhaps impossible. Extracting spatial information is further hampered by the limited temporal resolution of the sonar system (Simmons et al., 1989; Wiegrebe & Schmidt, 1996; Surlykke & Bojesen, 1996).

We have suggested before that limitations of the bat's sonar system result in bats having access to a sparse, low dimensional description of the outside 3D world only (Vanderelst et al., 2015, 2016). Some recent evidence seems to confirm this. First, Geberl et al. (2019) confirmed that biosonar has a low spatial resolution, both for distance and angular dimensions. Also, neurophysiological data suggest that closely spaced objects are not perceived seperately by echolocating bats. In contrast, neural responses to echo cascades from densely spaced objects seem to represent a single extended stimulus (Warnecke et al., 2018; Knowles et al., 2015).

However, it should be noted that such low dimensional descriptions have been shown to potentially contain a lot of useful information (Kuc, 1997b,a). We previously demonstrated that bats might use sonar template-based scene recognition (Vanderelst et al., 2016; Vanderelst & Peremans, 2017). Under this hypothesis, bats compare the acoustic signature of the environment to a set of previously stored signatures. We operationalized these acoustic signatures as time-intensity profiles, which could be derived directly from the cochlear output.

In this paper, we directly evaluate the hypothesis that sonar provides bats inherently with a low dimensional, yet functionally sufficient, description of the world. We test this hypothesis in two steps. In a first step, we generate an extensive collection of synthetic echo sequences. Next, we process the echo sequences using a model of the bat's auditory periphery (Wiegrebe, 2008). We assess the compressibility of the output of the cochlear model using Principal Component Analysis (PCA). The synthetic echo sequences differ greatly in complexity. Nevertheless, the output of the cochlear model can be mapped onto a low-dimensional space. In a second step, we show that echoes collected in real bat habitats can be mapped into the same low-dimensional space, without significant loss of information. We verify this by demonstrating that the

low dimensional feature space allows recognizing locations in real habitats.

Together these results suggest that echolocating bats live in a low dimensional perceptual world. The physics of the sonar system and the auditory periphery map the complexity of the 3D environment onto a small number of uncorrelated features, irrespective of whether we use artificial or real echo sequences. Nevertheless, this limited set of features is sufficiently rich for bats to recognize locations, and as we discuss, could also support flight control.

2. Methods

2.1. Synthetic echoes

In this section of the paper, we asses how the filtering in the auditory periphery reduces the complexity of input signals. We do this by presenting a simple generative model for echo sequences received by bats. Next, we assess whether the cochlear representation derived from these synthetic echo sequences can be projected into a low dimensional space.

We created artificial echo sequences by randomly spacing ideal point reflectors. The distances of the point reflectors were selected using a two-step process. First we selected n_s seed points s, uniformly distributed over a range from 1 to 7 meters. Next, for every seed point s, we selected n_c cloud points normally distributed around s, with standard deviation 0.25 m. We added a single reflector at a distance between 1 and 2 to each collection of points. This two step process to modeling natural reflectors was inspired by the results of Yovel et al. (2009), who found that echoes returning from vegetation could be better modeled using a similar two-step process than simply using uniformly spaced points. We generated 10 sets of points for each combination of $n_s \in [5, 10, 20, 50, 100, 200]$ and $n_c \in [5, 10, 20, 50, 100, 200]$. Synthetic echoes consisting of few reflectors are likely to consist of only weak echoes.

Each set of points was converted into an environmental impulse response with a sampling frequency f_s of 219 kHz. The environmental impulse response length was set to 7499 samples. These parameters correspond to the recording length and the sampling frequency used in the acoustic data collection (see below). In representing each point reflector by a Dirac impulse, we modeled attenuation of the echoes by introducing an amplitude calculated based on spherical spreading (both ways). The final amplitude of each Dirac impulse was determined by sampling from a normal distribution centered on the calculated amplitude and with a standard deviation of 10 dB. This final operation represents a very simple model of varying reflector strength due e.g., to shape and orientation of the reflecting feature.

The impulse responses were converted to echo sequences by convolving them with a bat-like emission. We used a 1 millisecond long linear frequency-swept cosine starting at 100 kHz and ending at 40 Khz. Again, these parameters

correspond to those used in the acoustic data collection described below. The echo sequences were dechirped as follows,

$$E_d = \mathcal{F}^{-1}((\mathcal{F}(S))^* . \mathcal{F}(E)) \tag{1}$$

with $\mathcal{F}$ denoting the Fourier transform and $()^*$ the complex conjugate. The symbols S, E, E_d refer to the emission, the echo sequence and the dechirped echo sequence, respectively. Examples of the resulting echo sequences are shown in figures 1 and 2.

Artificial echo sequences were filtered using a model of the bat's auditory periphery, as proposed by Wiegrebe (2008). This model simulates the temporal and spectral resolution of the bat's cochlea. The model consists of a gammatone filterbank followed by half-wave rectification and non-linear compression with an exponent of 0.4. Finally, the model includes a 1 kHz low-pass filter modeling the response characteristics of the inner hair cells. In the current implementation, we used a gammatone filterbank with 8 filters, centered around 70 kHz. The filters were spaced to correspond to equivalent rectangular bandwidths (Glasberg & Moore, 1990). The output of each channel was attenuated with a time (distance) weighted factor corresponding to the atmospheric attenuation at the channel's center frequency. This operation approximated frequency-dependent atmospheric attenuation. The first 6 ms of data were removed. This was done to ensure the data corresponded with the measurement data. As indicated below, we removed the first 6 ms of the measurement data to avoid including the pickup of the emitted signal.

Figure 1a depicts a single example of a synthetic echo sequence. It also illustrates the resulting response for each of the 8 channels of the Wiegrebe (2008) model (panels b-i). From these responses, it can be seen that they are highly correlated across channels. This is confirmed when analyzing the average cross-channel correlation for all synthetic echo sequences (Fig. 1i). We calculated the correlation $\bar{\rho}$ between the output X_i of each frequency channel i and the mean across channels $\bar{X}$ for each synthetic echo, as follows,

$$\bar{\rho} = \frac{1}{n} \sum_i \frac{\text{cov}(X_i, \bar{X})}{\sigma_{X_i}, \sigma_{\bar{X}}} \tag{2}$$

The distribution of $\bar{\rho}$ is plotted in 1j. From this distribution, it can be seen that the value of $\bar{\rho}$ ranged from 0.97 to > 0.99. From these high cross-channel correlations, we conclude that the output of the cochlear model can be summarized by averaging across channels without extensive loss of information.

An example of such a cross-channel sum is plotted in figure 1k. In correspondence to our previous work (Vanderelst et al., 2016), we refer to this average across frequency channels as a template.

We used principal components analysis (PCA) to assess the dimensionality of the templates generated from the synthetic echo sequences. We found that almost all of the variance ($> 99\%$) in the templates could be captured by 35 components (see figure 1l). Some examples
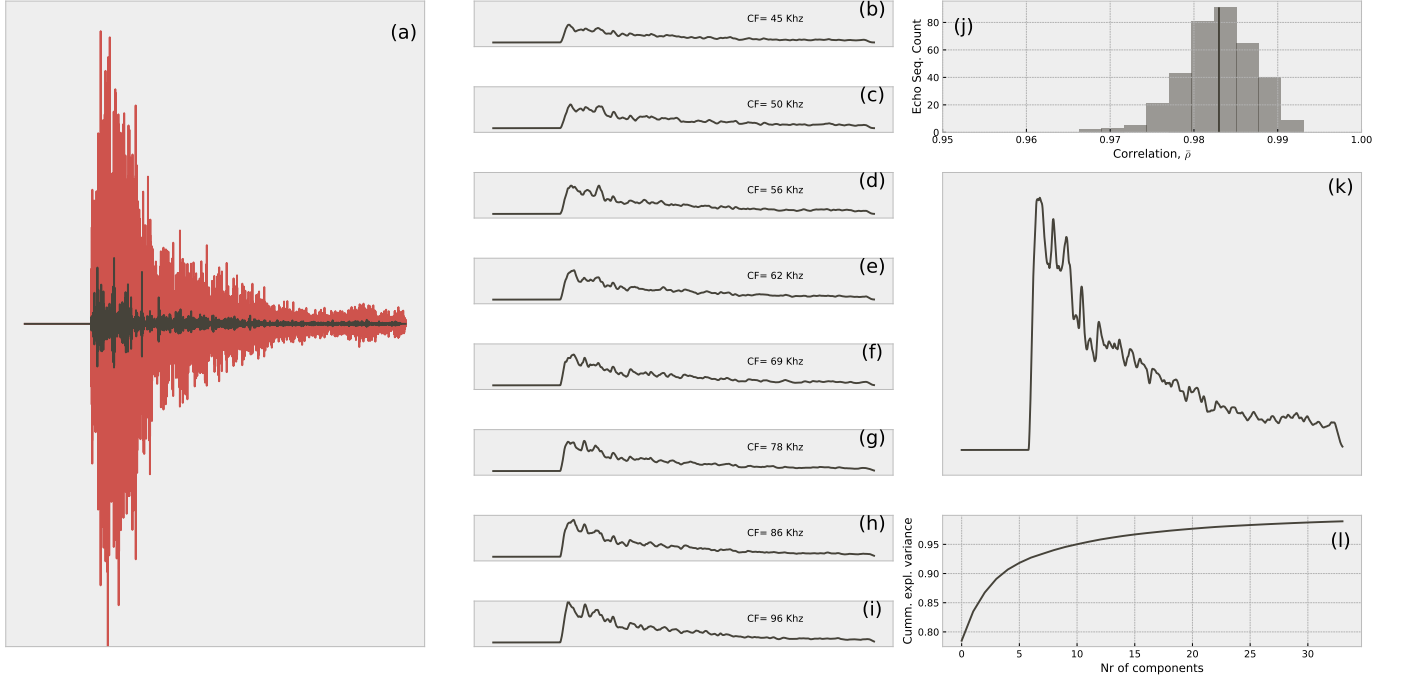
Figure 1: Example of synthetic echo sequence and its processing using the Wiegrebe (2008) model. (a) Example of a single synthetic echo sequence (red). The dechirped version is plotted in black. (b-i) Response of the 8 frequency channels in the Wiegrebe (2008) model. From these plots, it can be seen that channels correlate highly. This is confirmed in panel j, depicting the average correlation across channels for all synthetic echo sequences analyzed. Note that the x-scale does not start at zero. (k) The output of the Wiegrebe (2008) model summed across the frequency channels, is referred to as a template in this paper. Panel (l) shows the cumulative proportion of explained variance as a function of number of PC across all templates for all synthetic echo sequences.

of templates reconstructed from the principal components are shown in figure 2).

In this section, we used a generative model for naturalistic echoes. We assumed that echoes sequences received by bats can be considered as generated by anisotropically distributed point reflectors. Using this model, we have derived a low dimensional space ($D = 35$), which can be used to describe the templates resulting from the artificial echo sequences. This implies that, after filtering through (a model of) the bat's auditory system, these artificial echoes can be converted into low dimensional descriptions. The original echo sequences, generated by 25 to 40,000 point reflectors effectively map to a 35 dimensional vector. In the remainder of the paper, we will assess whether this low dimensional space allows describing the echoes from real bat habitats as collected using an ensonification device.

2.2. Natural Templates
2.2.1. Acoustic data

We reused the large library of echoes gathered in real bat habitats by Vanderelst et al. (2016). Below, we briefly discuss the construction of templates based on these echoes, and we refer to our previous work for more details. Vanderelst et al. (2016) constructed an ensonification device featuring 31 Knowles FG series microphones connected to a custom-built sonar data acquisition board (Steckel et al., 2013). The emitter was a Sensecomp 7000 Series ultrasonic broadband speaker. The ensonification device was mounted on a pan-tilt system allowing to rotate it in azimuth and elevation. The device and pan-tilt system was mounted on a tripod.

Here we use data collected along two linear pathways through natural habitats (Vanderelst et al., 2016). At the first site, a public park in Sde Boker, Israel (Fig. 3). The environment consisted of a walkway lined with rocks and vegetation. Unpublished bat monitoring data at the site confirmed that bats used this corridor to commute. The ensonification device was placed at 50 positions along a straight line spanning a total of 10 meters Locations were spaced 20 cm apart. At each location, echoes were collected for 31 azimuth directions (-150 to +150 degrees) and 7 elevation directions (-40 to +25 degrees). Therefore, at each location, echoes were collected from 31 x 7 = 217 directions. At each azimuth and elevation direction, three repeat measurements were recorded. In this way, a total of 32,550 echo trains (50 positions x 217 directions x 3 repeats) for each of the 31 microphones in the device were collected.

The second site, at the University of Bristol Royal Fort Gardens (UK), data were collected at 40 different positions along a line (Fig. 3). Locations were spaced 25 cm apart. As for the Israel site, we collected data for 217 directions, and measurements were repeated three times. This resulted in 26,040 echo trains (40 locations x 217 directions x 3 repeats) for each of 31 microphones. The recording of
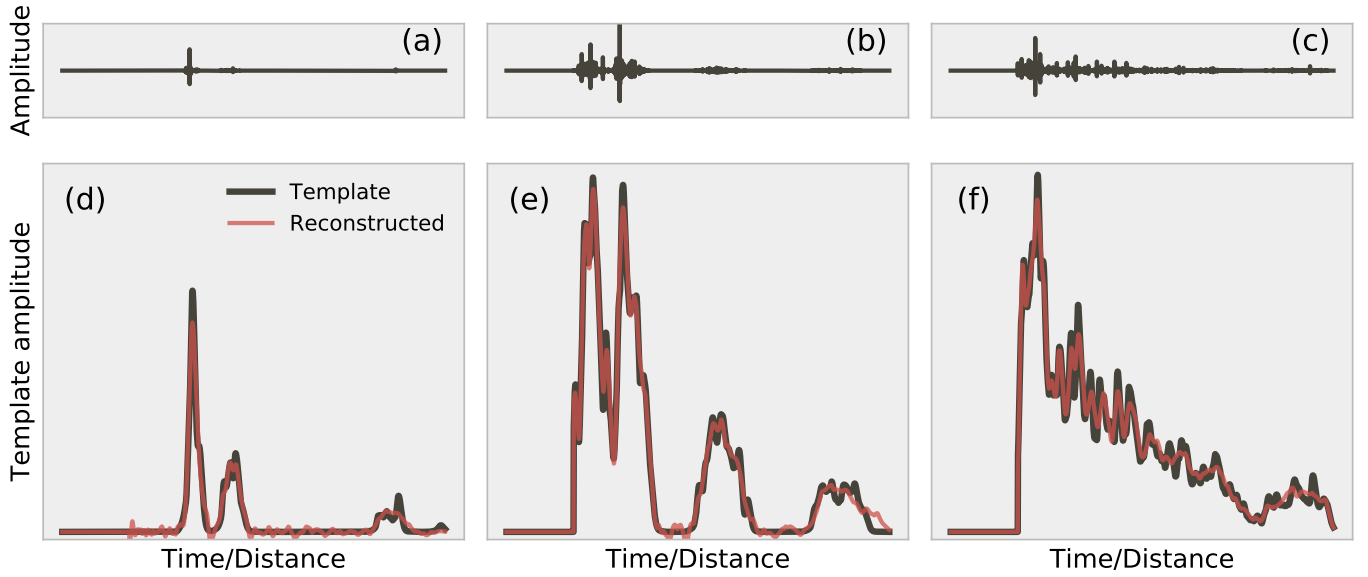
3

Figure 2: Example of synthetic dechirped echo sequences, their corresponding templates and the templates reconstructed from the 35 PCs. (xx-xx) Three examples of synthetic echo sequence waveforms. Panels xx-xx depict the resulting templates (i.e., summation across the output channels of the Wiegrebe (2008) model). These panels also show the result of projecting the templates into a 35D PC space and transforming the result back to the original coordinates.

the echoes started with the emission of a hyperbolic frequency modulated sweep and was recorded for 34 ms ($\sim$5.7 m) at a sampling rate of 219 kHz.

2.2.2. Template construction

For every azimuth and elevation direction at each of the 40 or 50 locations, a biologically plausible acoustic signature of the environment was derived from the echoes, i.e., a template. Below, we briefly outline the steps used to derive templates from the acoustic data.

As we did for the synthetic echo sequences, the echo trains collected in bat habitats were dechirped and filtered using a model of the bat's auditory periphery as proposed by Wiegrebe (2008). We concluded from analyzing the synthetic echoes that the response in different frequency channels is highly correlated. Therefore, the output of the cochlear model was averaged across frequency channels to obtain a time-intensity profile, or template. As the beam of the emitter used was more narrow than the typically combined hearing and emission directionality of bats (Vanderelst et al., 2010; Reijniers et al., 2010; Jakobsen et al., 2012) templates were averaged across three neighboring directions. The first 6 ms of data were removed to avoid using the emissions themselves for analysis. The processing steps outlined above resulted in 3 × 217 templates, for each location at each site, corresponding to 217 directions and three repeats.

Among the set of templates, some had very little energy in them, notably those collected for directions without reflectors. For example, aiming the ensonification device upwards often resulted in no echoes returning. These templates could not be classified (Vanderelst et al., 2016).

Therefore, we removed them from the current analyses by omitting all templates with very low energy. The resulting number of templates included in the present analyses are given in table 4.2.

2.2.3. Noise floor

Internal (and, potentially, external) noise results in template values greater than zero, even in the absence of reflectors returning echoes. To avoid training the neural network on this noise, we estimated the noise floor in our data by collecting templates while pointing the device to empty space. The maximum template value obtained from these template measurements was used as the noise floor $\mathcal{N}_f$. The value of the noise floor $\mathcal{N}_f$ used had a numerical value of 0.15. Any template values below this value were set to this value. As we argued before (Vanderelst et al., 2016), using the empirical measurement noise to estimate $\mathcal{N}_f$ is biologically warranted as bats have a better signal to noise ratio in their auditory system compared to our measuring device. Indeed, their dynamic range is at least 30 dB larger than our measurement device.

2.2.4. Mapping templates into the 35-dimensional space

The templates from both sites were mapped into the 35-dimensional space derived from the synthetic echoes, described above. We assessed how well the templates collected in real habitats can be represented by the 35 components by mapping the templates into this space and transforming the results back to the original space. Next, we calculated the correlation (and explained variance) between the original templates and the reconstructed templates.
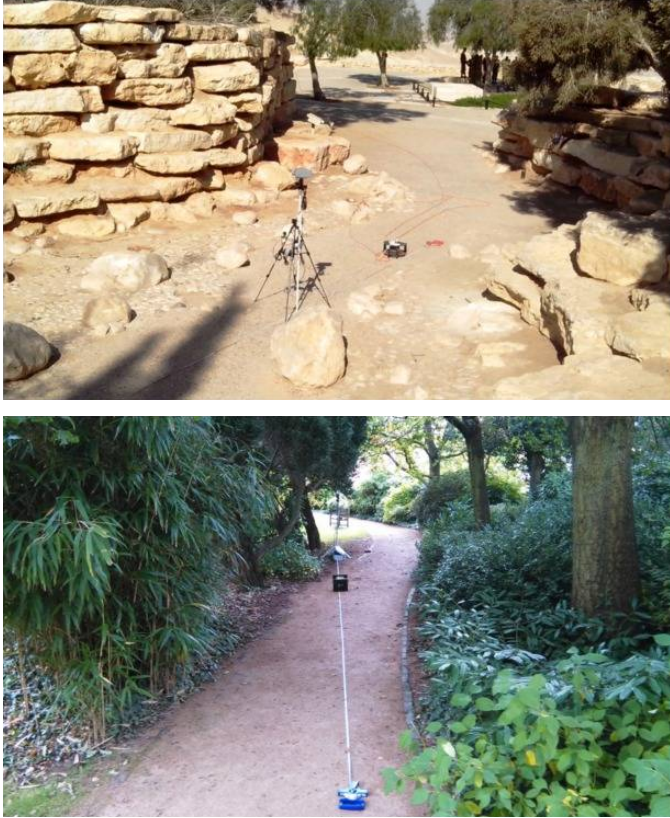
4

Figure 3: Pictures of the sonar data acquisition process. Top: the ensonification device used at the Israeli site at Sde Boker. Bottom: The device used at the Bristol University Royal Fort Gardens. The tape measure indicated the transect along which the device was positioned in 40 steps.

2.3. Artificial Neural Networks

We assessed whether the templates projected into the 35-dimensional space still contained sufficient information to be recognized. Previously, we assessed whether templates can be used to recognize locations using, what Baddeley et al. (2012) called, a perfect memory approach. Templates were stored in a lookup table. The Euclidean distance between two echo templates was the metric used to assess how well bats should be able to distinguish between templates. This table lookup and distance calculation requires a comprehensive search of memory that scales with the number of templates memorized. Hence, in our previous work, we established the theoretical possibility of classifying templates based on real echoes from environments bats were known to frequent, irrespective of the significant amount of memory and computational resources required to do so. Here, we use a more biologically plausible memory system: an artificial neural network, i.e. a neurally inspired way of distributed information storage (McLeod et al., 1998). While we acknowledge that current neural network training methods most likely have no biological counterpart, we propose that this approach provides us with a more biologically plausible appraisal of the memory and computational effort required to implement

a similar scheme in a bat brain.

Simple artificial feedforward neural networks were implemented using Tensorflow 2 for Python. The first layer of the neural network added Gaussian noise to each input sample ($\sigma(\mathcal{N}_s) = 0.01$). The following layers were fully connected and used the rectified linear activation function. The output layer uses a softmax activation function, i.e. the normalized exponential function. The loss was calculated as the cross-entropy between the target output and the actual output of the neural network. Weights were updated after each epoch using the Adam algorithm (Kingma & Ba, 2014) .

For each site, we trained three neural networks, each taking the 35-sample templates as input. The first neural network was trained to output the azimuth direction associated with the template (31 targets). The second network was trained to return the elevation (7 different targets). The final network returned the location (50 or 40 different targets). The targets for the network were encoded using the one-hot encoding scheme.

Training a different neural network for each of these features ensured that the training algorithm paid equal attention to each feature. When training a single network for all features, the optimization procedure would preferably reduce errors in the location predictions as this feature has the largest variance. The full architecture can be considered as a single network with different streams for each feature (fig. 4).

We should point out that we did not take any steps to avoid so-called overfitting by the neural networks. Overfitting occurs when a neural network (Ghotra & Dua, 2017), or other mathematical models (Hawkins, 2004), extract complicated relationships between inputs and outputs that do not hold beyond the training examples. Overfitting typically reduces the ability of the model to generalize its predictive power to new data. However, in this case, the model is not required to generalize to other environments. Instead, we are only interested in whether the neural networks can learn and classify the set of training examples. Nevertheless, it would be beneficial if the neural networks would be capable of classifying templates corresponding to directions that are slightly different than those included in the data sets. Indeed, this is a feature of the Euclidean distance measure used in implementing the perfect memory. In that approach, novel templates will be classified to positions with similar templates. We verified whether this properties was also found in the neural networks the case by generating a set of templates by interpolating in azimuth and elevation between the directions ensonified in the data set. In our data sets, we ensonified 31 different azimuth directions and 7 elevation directions. The interpolated data consisted of 30 azimuths and 7 elevations. We used the neural networks trained on the original data to classify these interpolated data.
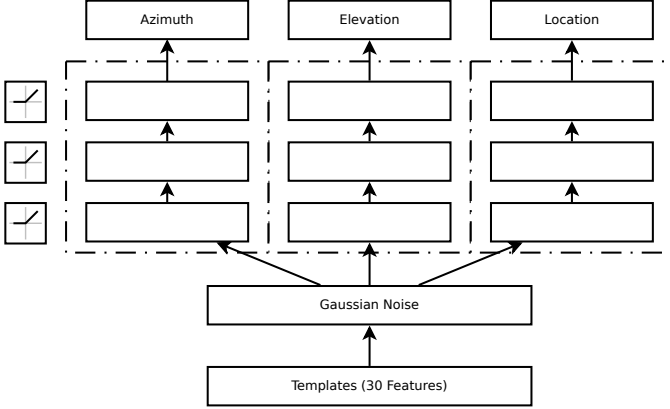
Figure 4: Depiction of the neural network architecture trained for each site.

2.3.1. Training examples

We wished to ensure that the neural networks encoded the templates and were able to recognize them in the presence of realistic noise. Indeed, it can be assumed that neural noise interferes with encoding or memorizing templates. Therefore, we trained the neural networks using templates to which noise was added.

In our previous work, we describe a procedure, based on Dau et al. (1996), to derive a realistic model for stochastic noise. In brief, we empirically determined a noise level that resulted in a 75% correct discrimination of two artificial bat calls differing by 2 dB in amplitude. This criterion is a conservative estimate of the just noticeable difference in loudness in bats. Based on this procedure we assumed internal (independent) Gaussian stochastic noise $\mathcal{N}_s$ for each sample with $\sigma(\mathcal{N}_s) = 0.01$ and $\mu = 0$.

2.4. Perfect Memory

To be able to directly compare the results of the neural networks with the perfect memory approach (Baddeley et al., 2012), we implemented a version of the classification scheme employed in our earlier work (Vanderelst et al., 2016). We generated noisy templates by adding Gaussian noise to them ($\sigma(\mathcal{N}_s) = 0.01$). Next, we performed the classification of these noisy templates by finding the closest template (according to the Euclidean distance) in the original set of templates. The location, azimuth, and elevation of the closest template were considered as the noisy template's classification. This was repeated ten times, and we report on the average performance across these repeats.

3. Results

We found that the templates derived from acoustic data collected in real habitats could be well represented by the 35 components derived from synthetic data. This 35-dimensional space could represent 98% of the variance in the templates. Figure 5 shows examples of templates and their reconstructed counterparts. Therefore, although the model used to generate echo sequences was very simple, these results indicate that the 35D feature space derived from the generative echo model could adequately represent templates resulting from real acoustic data.

In our previous work (Vanderelst et al., 2016), we showed that the templates could be used to recognize locations in bat habitats. As we have shown that the templates can be projected into a lower 35D feature space without loss of information, it can be expected that these 35 features can be used to recognize locations as well. We found this to be indeed the case.

Figure A.9 depicts the training history of all networks. This figure shows that, in all cases, training almost reached an asymptote around or before 250 epochs. However, some additional training can be seen up to 1000 epochs (when learning was arbitrarily halted).

Figure 6 summarizes the performance of the three neural networks trained on the Israel templates. The performance was obtained by assuming that the most active output neuron represented the classification of the network. Inspecting the cumulative distribution of error reveals that the neural network encoding performed as least as well as the perfect memory approach. The number of small errors was larger for the neural network than for the perfect memory. The neural network was able to classify about 80% of templates with an error of fewer than 30 degrees in azimuth. In elevation, almost all templates were classified with an error below 15 degrees. Most (78%) errors in location attribution were smaller than 0.5 meters. The confusion matrices confirm that most classification errors were small.

Interestingly, for azimuth and elevation, classification errors were not uniform across directions. This can be seen by the classification entropy plotted in figure 6. In particular, these curves show peaks around zero azimuth and medium elevations. The geometry of the environment can explain these peaks. The corridor (fig. 3) does not return a lot of echoes from azimuths around zero. Indeed, this direction is mostly open space. The same is true for elevations above 0.

Figure 7 shows the results for the Royal Fort Park data. These results mirror those in figure 6. Again the neural network performed at least as good as the perfect memory approach. For azimuth, the perfect memory approach slightly outperform the neural network (notice that y-scale in the top panels of this figure differ from figure 6). Overall, the performance for this data set was better than for the Israel data. This is probably due to the fact that echoes were collected closer to reflectors (vegetation) than in the Israel data set, which leads to strong echoes.

4. Discussion

4.1. The low dimensional world of bats

Reijniers & Peremans (2010) quantified the transmission of auditory information in the cochlea of bats. In
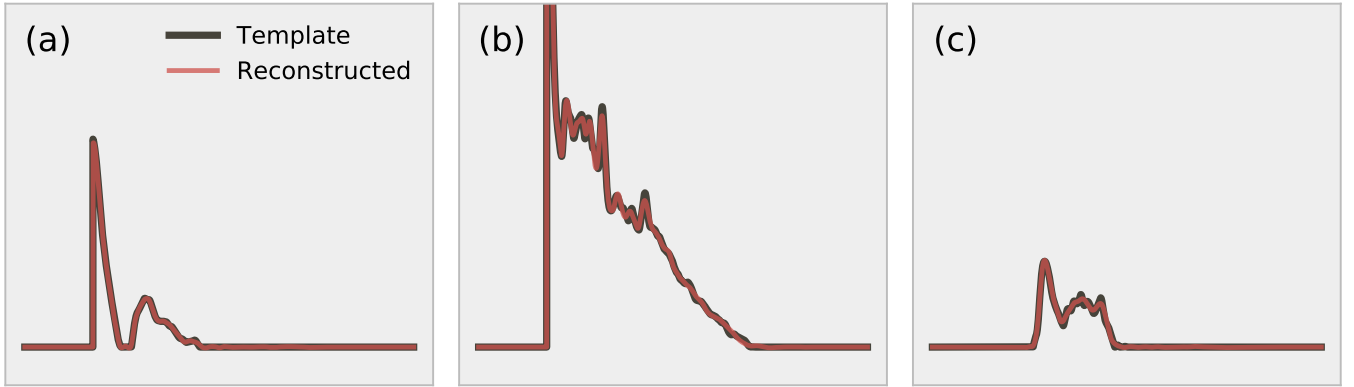
Figure 5: Examples of three templates generated from the echoes collected at the Israel and Royal Fort Gardens site. The panels also depict the result of projecting the templates in the 35-dimensional space and transforming the result back to the original coordinates.

that study, the authors adapted the Meddis inner hair cell model (Meddis, 2006) to assess the amount of information in the acoustic signal preserved by the cochlea. The authors concluded that spectral notches introduced by two overlapping echoes are faithfully encoded by auditory nerve spike patterns. However, quantifying the information transfer rate, they also found that about half of the information (in terms of bits) are lost in the transduction from acoustic energy to spike trains. In the current paper, we expand on this work and quantify the dimension reduction due to cochlear encoding. Moreover, we show that this reduction in complexity still supports an ecologically valid task: sonar-based scene recognition.

We assumed echoes received by bats can be modeled as originating from a large collection of point reflectors. Passing these synthetic echoes through a model of the auditory periphery, we obtained a time-intensity profile (template) for each echo sequence. These templates could be well described by a low dimensional space of 35 principal components (figure 2). We found that the same 35D space could be used to describe templates originating from echoes collected in real bat habitats (figure 5). This suggests that the filtering of the auditory periphery provides the bat with a significantly compressed acoustic description of its environment. To identify the locus of the compression, we performed a number of additional analyses. In particular, we evaluated whether the low-dimensional description is a property of the echo sequences themselves, or whether this property results from the finite bandwidths of the gammatone filters and/or the low-pass filter in the Wiegrebe (2008) model.

We generated templates from the synthetic echo sequences using two versions of the Wiegrebe (2008) model; with and without the 1 Khz low-pass filter. Per default, the model includes this filter to model the response characteristics of the inner hair cells (Meddis, 2006; Reijniers & Peremans, 2010). By removing this lowpass filter, we can assess to what extent it is responsible for the observed compressibility of the templates. We also generated templates using echo sequences of which the samples were randomly shuffled. This allows inferring whether the low-dimensionality of the templates is a property of the echo signals (before being filtered by the cochlear model).

Figure 8 shows the number of PCs needed to capture 99% of the templates variance. From this plot, it can be seen that the number of PCs increases when shuffling the echo sequences from 35 to 63Ȯmitting the 1 kHz filter from the Wiegrebe (2008) model increases the number of PCs to 252. Including the 1 kHz filter, the average spectrum of the templates (fig. 8, inset) shows that most energy in the templates is contained in frequencies below 1 kHz. Therefore, if we take 1 kHz as the highest relevant frequency, it ought to be possible to completely represent templates by sampling them at 2 kSamples/sec. This corresponds to about 70 samples per template, as templates consist of 7499 samples at 219 kHz sample rate. This corresponds roughly to the number of PCs required to represent templates derived from echo sequences that were shuffled (fig. 8). The number of PCs required to represent unshuffled templates is about half of this, i.c., 35. Therefore, we conclude that the 1 kHz filter is the main, but not exclusive, locus of the dimensionality reduction. The structure of the echo signals also decreases the dimensionality of the templates. The sonar system of bats, through a number of filtering stages, compresses the 3D complexity of the world onto a lower dimensional representation.

Despite the low dimensional character of the templates, they are sufficient to recognize environments. Indeed, this was already demonstrated in our previous work (Vanderelst et al., 2016). Here, we directly demonstrated that the 35 principal components are sufficient to discriminate the associated locations and viewing directions using neural networks. The neural networks' performance was somewhat better than the perfect memory approach we used previously (figs. 6, 7).

While recognizing location in the environment is a critical component of navigation, bats use their sonar not only for navigating. They also rely on sonar for flight control.
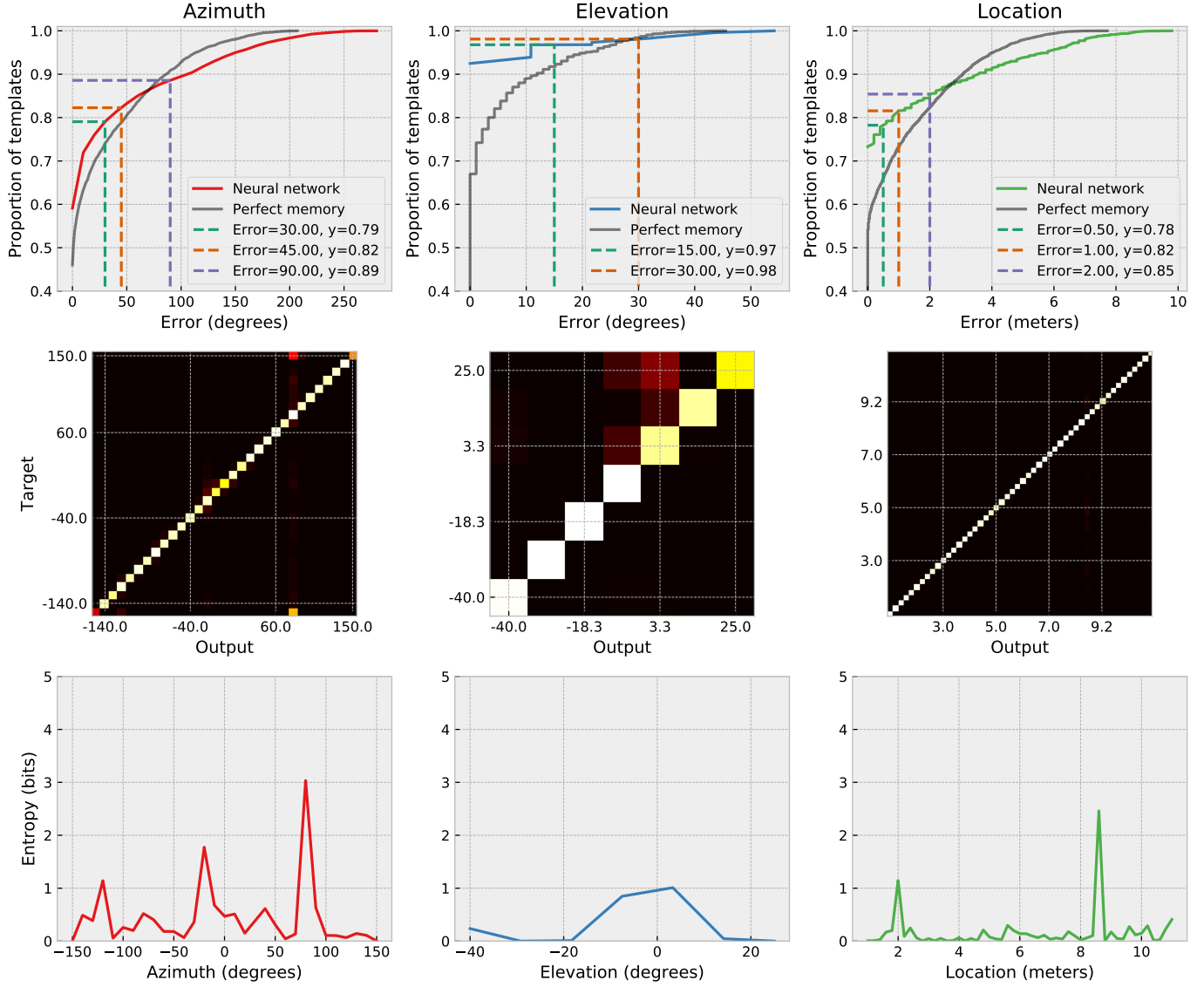
Figure 6: Classification results for the Israel site. Top row: cumulative error distributions for the azimuth, elevation and location classification. Middle row: Confusion matrices for the classifications. Bottom row: remaining classification entropy for each target (derived from the confusion matrices).
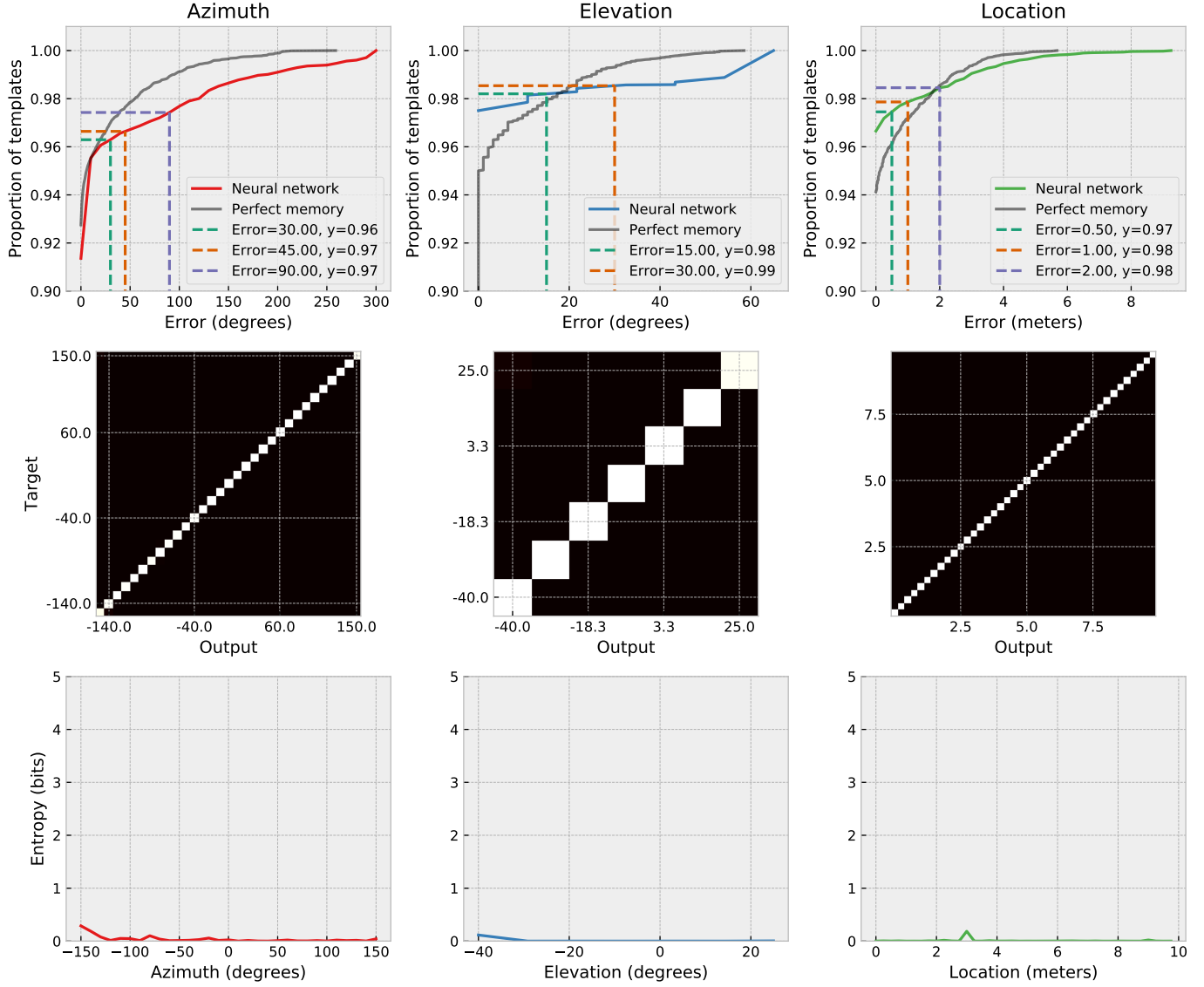
Figure 7: Classification results for the Royal Fort Park site. Top row: cumulative error distributions for the azimuth, elevation and location classification (Notice the different y-scale for these plots than in figure 6). Middle row: Confusion matrices for the classifications. Bottom row: remaining classification entropy for each target (derived from the confusion matrices).
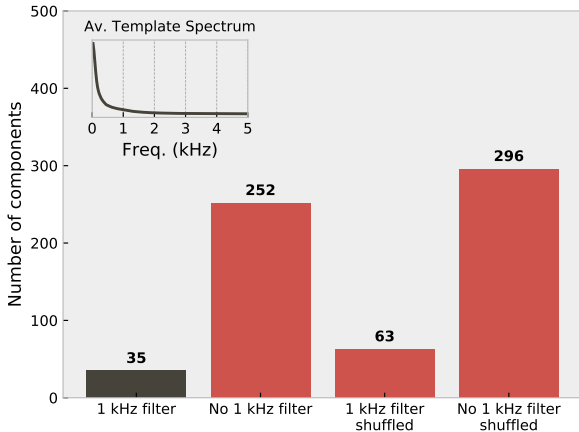
Figure 8: Barplot depicting the number of required PCs to capture > 99% of the variance in the templates derived from synthetic echoes. The label *1 kHz filter* refers to whether the 1 kHz filter was omitted in the implementation of the Wiegrebe (2008) model. The label *shuffled* refers to whether the samples of the echoes were randomly shuffled before generating the templates. The inset shows the average spectrum of the templates generated using non-shuffled echoes and a version of the Wiegrebe (2008) model including a 1 kHz filter (i.e., the default model).

This includes, avoiding obstacles and/or following guiding structures. The environmental description the bat has access to for flight control starts from the same cochlear output used for place recognition. Therefore, our current work suggests that the same low-dimensional description should be sufficient for flight control as well. Indeed, we have previously demonstrated that a low-dimensional description is sufficient for obstacle avoidance and corridor following (Vanderelst et al., 2015; Mansour et al., 2019). In that work, we converted both the echoes at the left and at the right ear into a template-like representation used as input to the motion controller. Subsequent work will have to demonstrate whether the parameters of the 35D space presented here can be used directly for controlling flight. In other words, we conjecture it should be possible to extract 35 features from an echo signal and convert this vector into a motor command. One could presume a neural network that converts the 35 features to flight control parameters.

4.2. Template encoding

We were unable to find values for the number of neurons in the cortex of bat species. However, the mouse cortex has been estimated to contain about $9 \cdot 10^4$ neurons or $7 \cdot 10^8$ synapses per cubic millimeter (Braitenberg & Schüz, 2013). The neural networks presented in this paper consisted of 250 neurons (not including the input and output layers) and 50 thousand connections. Therefore, to the extent that the proposed neural networks can be considered as biologically plausible models of memory, they are orders of magnitude smaller than a single cubic mil-

Site	T	n	W	W (nb)	R
Israel	183430	5395	52875	51975	28.8%
Royal	264146	7769	51345	50475	19.4%

Table 1: Table indicating the size of both template data sets and number of weights in the networks. The template size T is obtained as $35 \times n$, with n the number of templates in the set. The number of weights W includes all connections between neurons as well as the bias weights for each of the three networks trained. $W(nb)$ omits the bias weights. The ratio R gives the ratio W/T.

limeter of mouse cortex. Let alone the complete cortex, which has been estimated to contain 75 million neurons (Williams, 2000).

Despite their small size, the networks in this paper were able to encode (memorize) the templates at least as well as the perfect memory approach. The neural networks presented here are tiny compared to the density of cortical matter. However, the networks are also small compared to the data they encode. Table 4.2 lists the number of weights for the artificial neural networks and compares this number with the size of the 35-dimensional templates. The number of floating point numbers making up the templates is larger than the number of weights by a factor of 3-5. Therefore, the networks have a smaller memory footprint than the data they encode. It should be noted that we have not made an attempt to try and to optimize (minimize) the size of the networks. From these numbers, we conclude that the low-dimensional signals provided by the bat's sonar system result in a low memory footprint and that they would use little neural resources to be stored and processed.

4.3. Conclusion

It has been suggested before that bats' sonar systems provides detailed information about the world. A number of authors have referred to sonar as providing the bat with an acoustic image (Lee et al., 2017; Barchi et al., 2013; Moss & Surlykke, 2001; Schnitzler et al., 2003; Simmons, 2012; Ulanovsky & Moss, 2008; Clare & Holderied, 2015; Surlykke et al., 2016; Geipel et al., 2013). As mentioned above, this is at odds with some recent behavioral and neurophysiological findings in bats faced with complex echo sequences (Knowles et al., 2015; Geberl, 2013; Warnecke et al., 2018). However, it is also at odds with the various filtering stages involved in the emission and the reception of bio-sonar. In particular, we argued before that the neurophysiology of the bats auditory system represents a bottleneck in the amount of raw echo signal information accessible to the bat. Indeed, Reijniers & Peremans (2010) already noticed that significant compression occurs during cochlear transduction. Here, we followed up on that work to quantify the dimensional reduction of the input signals as a result of peripheral auditory processing.

From our current work, in combination with previous results (Vanderelst et al., 2016, 2015; Mansour et al., 2019;

Reijniers & Peremans, 2010), we conclude that the sonar system of bats maps complex echo signals onto a limited number of features. These features can be efficiently stored to recognize scenes. Moreover, we conjecture that a (binaural) version of the features should allow for sonar-based flight control. As such, a view of echolocating bats living in a low dimensional sensory space emerges that is at odds with the widespread notion of sonar as a sensing modality that provides detailed acoustic images.

References

Aytekin, M., Grassi, E., Sahota, M., & Moss, C. F. (2004). The bat head-related transfer function reveals binaural cues for sound localization in azimuth and elevation. *The Journal of the Acoustical Society of America*, *116*, 3594–3605.

Baddeley, B., Graham, P., Husbands, P., & Philippides, A. (2012). A model of ant route navigation driven by scene familiarity. *PLoS computational biology*, *8*, e1002336.

Barchi, J. R., Knowles, J. M., & Simmons, J. A. (2013). Spatial memory and stereotypy of flight paths by big brown bats in cluttered surroundings. *Journal of Experimental Biology*, *216*, 1053–1063.

Braitenberg, V., & Schüz, A. (2013). *Cortex: statistics and geometry of neuronal connectivity*. Springer Science & Business Media.

Clare, E. L., & Holderied, M. W. (2015). Acoustic shadows help gleaning bats find prey, but may be defeated by prey acoustic camouflage on rough surfaces. *Elife*, *4*, e07404.

Dau, T., Püschel, D., & Kohlrausch, A. (1996). A quantitative model of the "effective"signal processing in the auditory system. i. model structure. *The Journal of the Acoustical Society of America*, *99*, 3615–3622.

Davis, W. H., & Barbour, R. W. (1970). Homing in blinded bats (myotis sodalis). *Journal of Mammalogy*, *51*, 182–184.

Franz, M. O., & Mallot, H. A. (2000). Biomimetic robot navigation. *Robotics and autonomous Systems*, *30*, 133–153.

Geberl, C. (2013). *Spatial and temporal resolution of bat sonar*. Ph.D. thesis Ludwig-Maximilians-Universität München.

Geberl, C., Kugler, K., & Wiegrebe, L. (2019). The Spatial Resolution of Bat Biosonar Quantified with a Visual-Resolution Paradigm. *Current Biology*, (p. S0960982219304750).

Geipel, I., Jung, K., & Kalko, E. (2013). Perception of silent and motionless prey on vegetation by echolocation in the gleaning bat micronycteris microtis. *Proceedings. Biological sciences / The Royal Society*, *280*, 20122830.

Ghotra, M. S., & Dua, R. (2017). *Neural Network Programming with TensorFlow: Unleash the power of TensorFlow to train efficient neural networks*. Packt Publishing Ltd.

Glasberg, B. R., & Moore, B. C. (1990). Derivation of auditory filter shapes from notched-noise data. *Hearing research*, *47*, 103–138.

Hawkins, D. M. (2004). The Problem of Overfitting. *Journal of Chemical Information and Computer Sciences*, *44*, 1–12.

Jakobsen, L., Ratcliffe, J. M., & Surlykke, A. (2012). Convergent acoustic field of view in echolocating bats. *Nature*, *493*, 93–96.

Jensen, M. E., Moss, C. F., & Surlykke, A. (2005). Echolocating bats can use acoustic landmarks for spatial orientation. *Journal of experimental biology*, *208*, 4399–4410.

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, .

Knowles, J. M., Barchi, J. R., Gaudette, J. E., & Simmons, J. A. (2015). Effective biosonar echo-to-clutter rejection ratio in a complex dynamic scene. *The Journal of the Acoustical Society of America*, *138*, 1090–1101.

Kuc, R. (1997a). Biomimetic sonar recognizes objects using binaural information. *The Journal of the Acoustical Society of America*, *102*, 689–696.

Kuc, R. B. (1997b). Biomimetic sonar differentiates coin head from tail. *The Journal of the Acoustical Society of America*, *101*, 3198.

Lee, W.-J., Falk, B., Chiu, C., Krishnan, A., Arbour, J. H., & Moss, C. F. (2017). Tongue-driven sonar beam steering by a lingual-echolocating fruit bat. *PLoS biology*, *15*, e2003148.

Mansour, C. B., Koreman, E., Laurijssen, D., Steckel, J., Peremans, H., & Vanderelst, D. (2019). Robotic models of obstacle avoidance in bats. In *The 2018 Conference on Artificial Life: A Hybrid of the European Conference on Artificial Life (ECAL) and the International Conference on the Synthesis and Simulation of Living Systems (ALIFE)* (pp. 463–464). MIT Press.

McLeod, P., Plunkett, K., & Rolls, E. T. (1998). *Introduction to connectionist modelling of cognitive processes.*. Oxford University Press.

Meddis, R. (2006). Auditory-nerve first-spike latency and auditory absolute threshold: a computer model. *The Journal of the Acoustical Society of America*, *119*, 406–417.

Moss, C. F., & Surlykke, a. (2001). Auditory scene analysis by echolocation in bats. *The Journal of the Acoustical Society of America*, *110*, 2207–2226.

Mueller, H., & Emlen, J. (1957). Homing in bats. *Science*, *126*, 307–308.

Müller, R., & Kuc, R. (2000). Foliage echoes: a probe into the ecological acoustics of bat echolocation. *The Journal of the Acoustical Society of America*, *108*, 836–845.

Reijniers, J., & Peremans, H. (2010). On population encoding and decoding of auditory information for bat echolocation. *Biological Cybernetics*, *102*, 311–326.

Reijniers, J., Vanderelst, D., & Peremans, H. (2010). Morphology-induced information transfer in bat sonar. *Physical review letters*, *105*, 148701.

Schnitzler, H.-U., Moss, C. F., & Denzinger, A. (2003). From spatial orientation to food acquisition in echolocating bats. *Trends in Ecology & Evolution*, *18*, 386–394.

Simmons, J. a. (2012). Bats use a neuronally implemented computational acoustic model to form sonar images. *Current Opinion in Neurobiology*, *22*, 311–319.

Simmons, J. A., Freedman, E. G., Stevenson, S. B., Chen, L., & Wohlgenant, T. J. (1989). Clutter interference and the integration time of echoes in the echolocating bat, Eptesicus fuscus. *The Journal of the Acoustical Society of America*, *86*, 1318–1332.

Steckel, J., Boen, A., & Peremans, H. (2013). Broadband 3-D sonar system using a sparse array for indoor navigation. *IEEE Transactions on Robotics*, *29*, 161–171.

Steckel, J., & Peremans, H. (2013). Batslam: Simultaneous localization and mapping using biomimetic sonar. *PloS one*, *8*, e54076.

Stones, R. C., & Branick, L. P. (1969). Use of hearing in homing by two species of myotis bats. *Journal of Mammalogy*, *50*, 157–160.

Surlykke, A., & Bojesen, O. (1996). Integration time for short broad band clicks in echolocating fm-bats (eptesicus fuscus). *Journal of Comparative Physiology A*, *178*, 235–241.

Surlykke, A., Simmons, J. A., & Moss, C. F. (2016). Perceiving the world through echolocation and vision. In *Bat Bioacoustics* (pp. 265–288). Springer.

Ulanovsky, N., & Moss, C. F. (2008). What the bat's voice tells the bat's brain. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 8491–8498.

Vanderelst, D. (2015). Social network analysis as a tool for research policy. *Plos Neglected Tropical Diseases*, . In press.

Vanderelst, D., De Mey, F., Peremans, H., Geipel, I., Kalko, E., & Firzlaff, U. (2010). What noseleaves do for fm bats depends on their degree of sensorial specialization. *PLOS ONE*, *5*, e11893.

Vanderelst, D., & Peremans, H. (2017). A computational model of mapping in echolocating bats. *Animal Behaviour*, *131*, 73–88.

Vanderelst, D., Peremans, H., & Holderied, M. (2015). Sensorimotor model of obstacle avoidance in echolocating bats. *PLoS Computational Biology*, . Paper accepted pending minor revisions.

Vanderelst, D., Steckel, J., Boen, A., Peremans, H., & Holderied, M. W. (2016). Place recognition using batlike sonar. *Elife*, *5*, e14188.

Verboom, B., Boonman, A. M., & Limpens, H. J. (1999). Acoustic perception of landscape elements by the pond bat (myotis dasycneme). *Journal of Zoology*, *248*, 59–66.

Von Helversen, O., & Winter, Y. (2005). Bat ecology. chapter Glossophagine Bats and Their Flowers: Costs and Benefits for Plants and Pollinators. (p. 346). University of Chicago Press.

Warnecke, M., Macías, S., Falk, B., & Moss, C. F. (2018). Echo interval and not echo intensity drives bat flight behavior in structured corridors. *The Journal of Experimental Biology*, *221*, jeb191155.

Wiegrebe, L. (2008). An autocorrelation model of bat sonar. *Biological cybernetics*, *98*, 587–595.

Wiegrebe, L., & Schmidt, S. (1996). Temporal integration in the echolocating bat, megaderma lyra. *Hearing research*, *102*, 35–42.

Williams, R. W. (2000). Mapping genes that modulate mouse brain development: a quantitative genetic approach. In *Mouse brain development* (pp. 21–49). Springer.

Williams, T. C., Williams, J. M., & Griffin, D. R. (1966). The homing ability of the neotropical bat phyllostomus hastatus, with evidence for visual orientation. *Animal behaviour*, *14*, 468–473.

Yovel, Y., Franz, M. O., Stilz, P., & Schnitzler, H. U. (2008). Plant classification from bat-like echolocation signals. *PLoS Computational Biology*, *4*.

Yovel, Y., Stilz, P., Franz, M. O., Boonman, A., & Schnitzler, H. U. (2009). What a plant sounds like: The statistics of vegetation echoes as received by echolocating bats. *PLoS Computational Biology*, *5*.

Yu, C., Luo, J., Wohlgemuth, M., & Moss, C. F. (2019). Echolocating bats inspect and discriminate landmark features to guide navigation. *The Journal of Experimental Biology*, (p. jeb.191965).

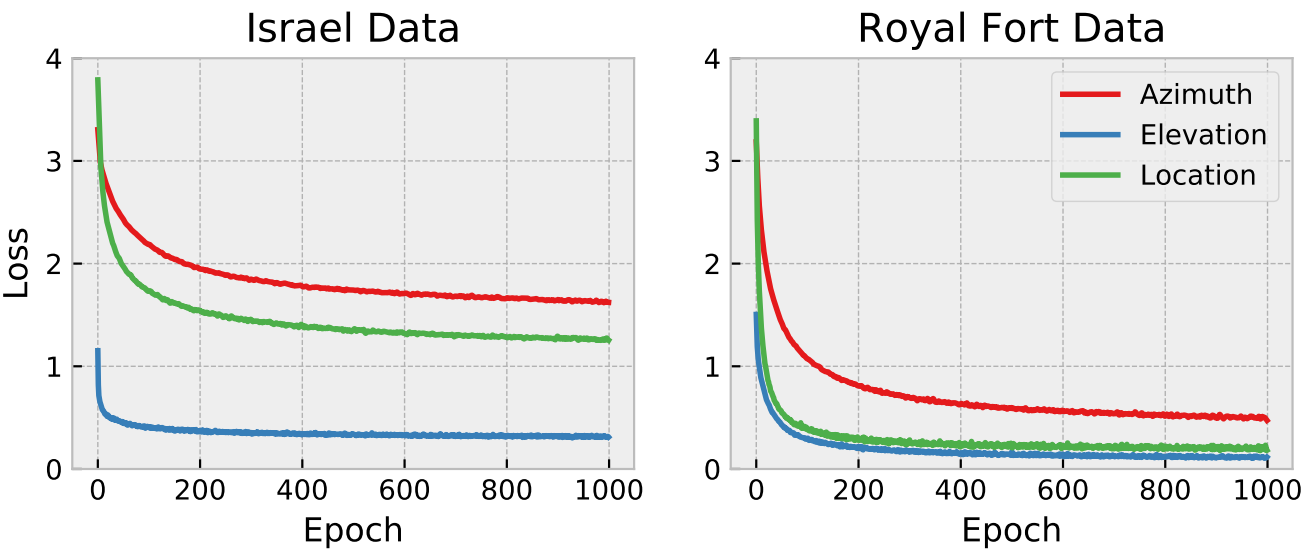**Appendix A. Supplementary material**

Figure A.9: Training history of the neural networks. The plots give the value of the loss function as a function of training epoch. Notice the logarithmic x-axis.
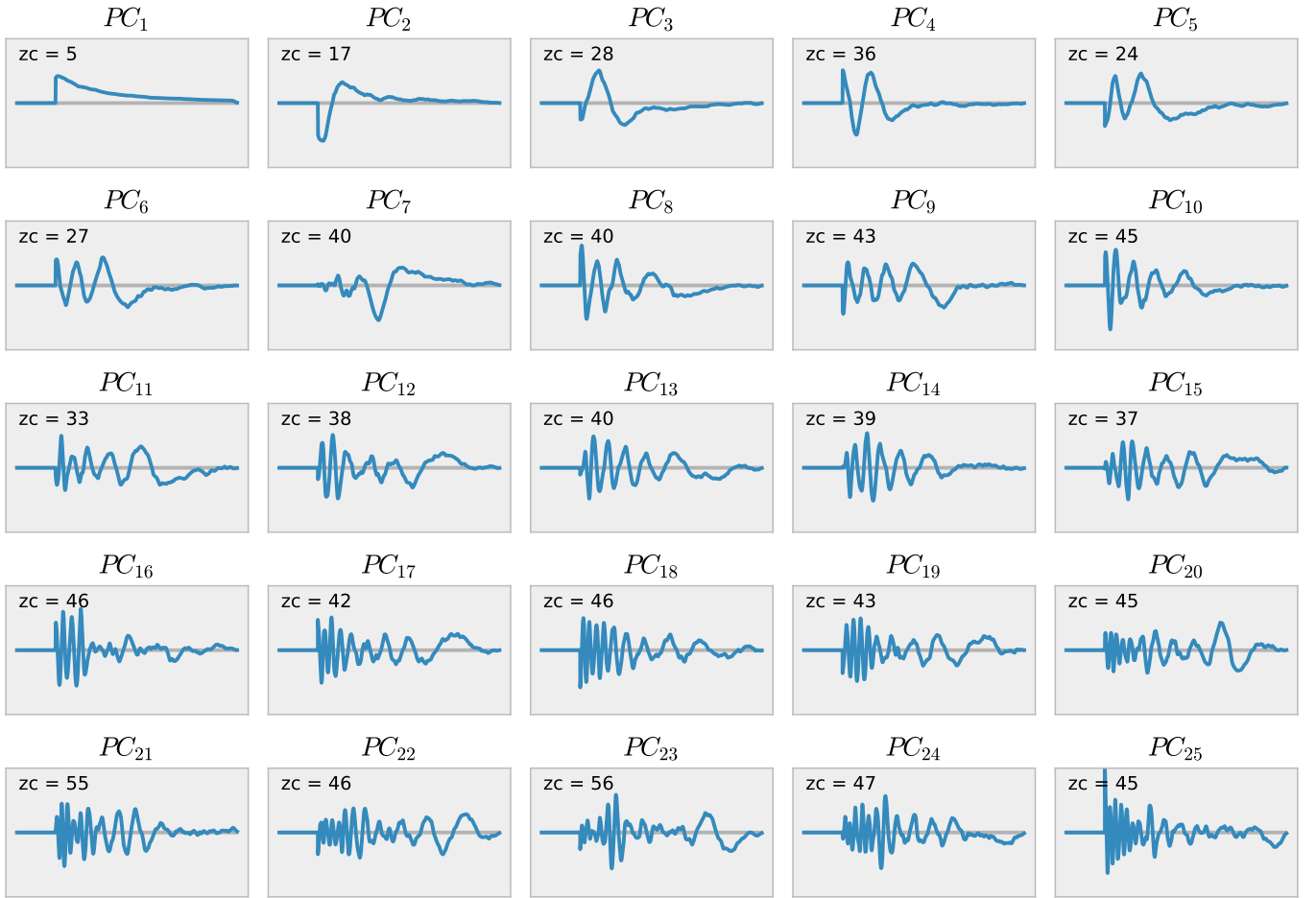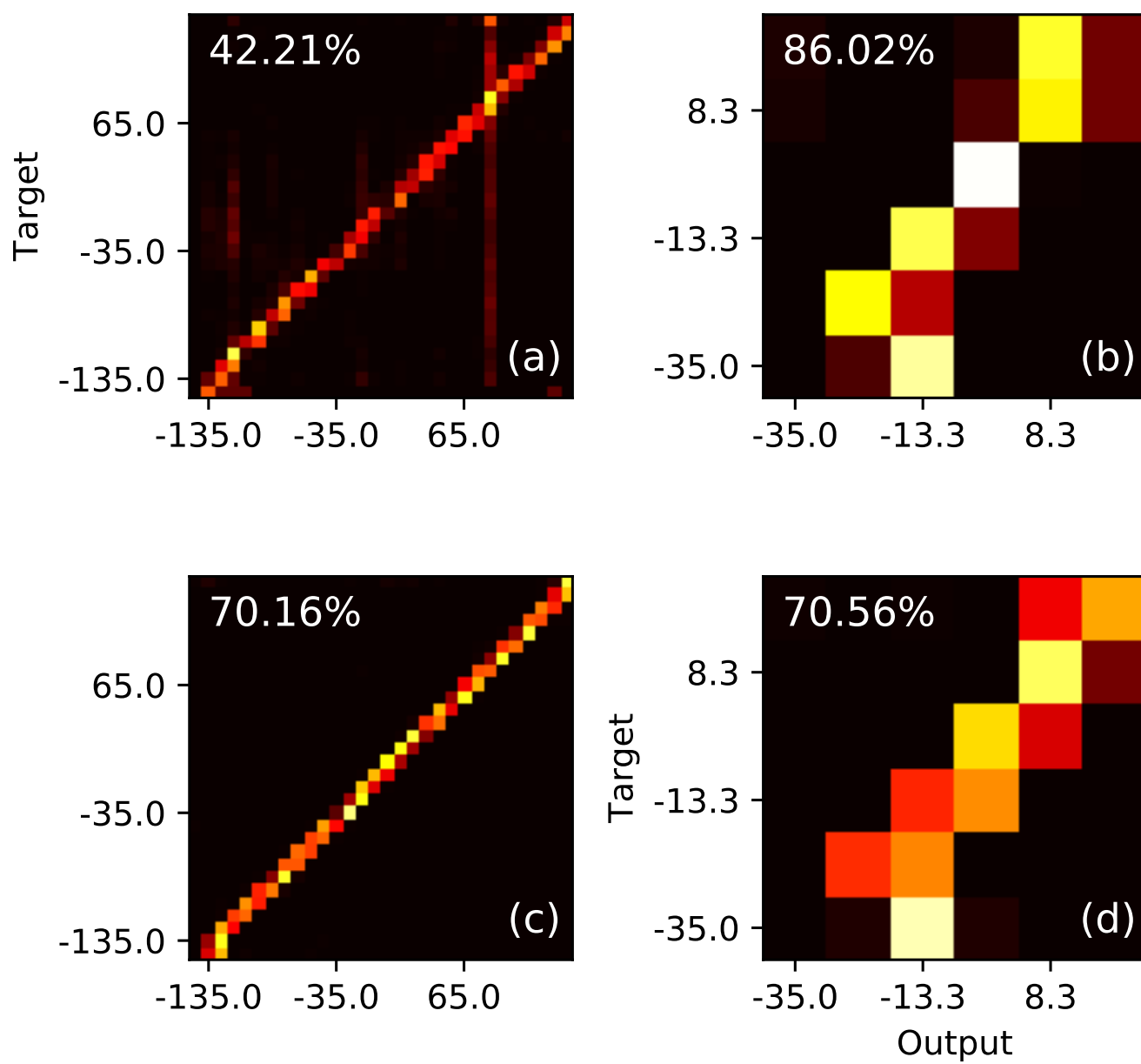
Figure A.10:

14

Figure A.11: sadasasasdasd