Animal Mimicry for Covert Communication with Arbitrary Output Distribution: Beyond the Assumption of Ignorance

Krzysztof Władysław ZUBER
Wroclaw University of Science and Technology, Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland, krzysztof.zuber@pwr.edu.pl

Krzysztof J. OPIELIŃSKI
Wroclaw University of Science and Technology, Wybrzeże Wyspiańskiego 27,
50-370 Wrocław, Poland, krzysztof.opielinski@pwr.edu.pl

Abstract

The paper describes a new method of embedding human communication in acoustic sequences mimicking animal communication. This is done to ensure a low probability of detection (LPD) transfer of covert messages. The proposed scheme mimics not only individual sounds, but also the imitated species' communication structure. This paper presents a step forward in animal communication mimicry – from pure vocal imitation without regard for the plausibility of communication's structure, through Zipf's law-preserving scheme, to the mimicry of a known communication structure. Unlike previous methods, the updated scheme does not rely on third parties' ignorance of the imitated species' communication structure beyond Zipf's law – instead, the new method enables one to encode information in a known zeroth-order Markov model. The paper describes a method of encoding an arbitrary message in a syntactically plausible, species-specific sequence of animal sounds through evolutionary means. A comparison with the previous iteration of the method is also presented.

Keywords: animal mimicry, covert communication, hidden Markov model

1. Introduction

Successful transmission of covert messages depends not only upon an appropriately robust coding method, but also upon decreasing the probability of their interception – for instance, in the case of underwater communication, the mere fact of a submarine sending a message can lead to its detection, regardless of whether the message is decoded or not. To limit the probability of detection of message transmission, a nature-inspired method of mimicry has been proposed [1]. This method relies on sending the hidden message as a string of an animal species' vocal signals.

However, even using animal sounds to transmit a message is not sufficient for a successful mimicry. This is because animals in general, and especially aquatic mammals (such as dolphins and whales), use complex structures in their communication [2,3,4]. In fact, any intelligent communication tends to distribute its units according to Zipf's law [3,5,6] which is not evident in superficial animal mimicry schemes. Thus, even a simple, automatic rank-frequency plot of received vocals will invalidate these schemes.

To address this problem, a mimicry method based upon generating a syntax for encoding covert messages, which gives the output units a Zipfian distribution while maximising both the information throughput and output distinctness was devised and described [7]. However, even that method has serious limitations – the most important being that it pre-supposes that third parties are ignorant of the mimicked species' communication structure beyond the very general fact that it obeys the Zipf's law.

Therefore, a new method of mimicry was designed with the aim of encoding and decoding arbitrary messages composed of symbols as sequences of animal vocal units with a specified distribution. The method relies on constructing hidden Markov models. The scheme has been tested and compared to the alternative, Zipfian method, and proved to be the better option when it comes to successful mimicry, while having comparable information throughput rate.

2. Hidden Markov Model

To achieve the desired operation of encoding – namely, the ability to encode data expressed as **symbols** with arbitrary frequencies as animal species' vocal units with a specified probability distribution (for instance, the probability distribution of syllables in birdsong [8] or of temporal spacing between dolphin pulses [9]) – the hidden Markov model (HMM) approach was chosen.

Hidden Markov models rely fundamentally on transitions within hidden, unobservable states, which can probabilistically generate **observations** or **outputs** [10]. Transitions between hidden states are described by the stochastic matrix A (called the **state transition matrix**), and observation generation is described by matrix B, called the **output(emission) matrix**, where $A_{ij} = P\left(s_{t+1} = x_j \middle| s_t = x_i\right)$, $B_{ij} = P\left(o_t = y_j \middle| s_t = x_i\right)$, $X = \{x_1, x_2, \cdots, x_N\}$ are the N possible hidden states (**state alphabet**), $Y = \{y_1, y_2, \cdots, y_M\}$ are the M possible observations (**output alphabet**), $S = \{s_1, s_2, \cdots, s_T\}, S \in X$ is the sequence of hidden states at step t out of T (**state sequence**), and $O = \{o_1, o_2, \cdots, o_T\}, o \in Y$ is the sequence of observations in T steps (**observed sequence**) $(N, M, T \in \mathbb{N}^+)$. In addition, the probability of the system being initially in any given hidden state (**initial state distribution**) is represented by $\pi = \{\pi_1, \pi_2, \cdots, \pi_N\}$, $\sum_{m=1}^{N} \pi_m = 1$.

For the purpose of encoding, the hidden states are assumed to correspond to the characters of the message to be sent, and the output states – the emitted vocal units (which, it must be remembered, do not necessarily correspond to vocal syllables, but can also represent separations between units in communication dependent on timing or the number of repetitions in communication dependent on unit repetition [2]).

For successful mimicry, the probability distribution of output units must follow the probability distribution of units in the mimicked species' vocal communication, $\forall t\, P(o_t = y_i) = F(i)$, $F(i) \in [0,1]$, $\sum_{i=1}^{M} F(i) = 1$. Similarly, for the input symbols, $\forall t\, P(s_t = x_i) = G(i)$, $G(i) \in [0,1]$, $\sum_{i=1}^{N} G(i) = 1$. Thus, it is assumed that both symbols' and observations' probabilities are independent.

A trivial solution of the problem of constructing A, B and π would be to set them to $A_{ij} = G(j)$, $B_{kl} = F(l)$, $\pi = \upsilon$, where υ is the left eigenvector of A with the eigenvalue of 1 (i.e. the stationary distribution of the Markov chain described by matrix A), which ensures that the system starts at steady state and does not need time to settle for the desired output distribution.

However, doing so, while ensuring the proper input and output distributions, prevents any encoded message from being decoded with any certainty. To understand why this is the case, and devise steps to be taken to remedy the problem, it is illustrative to consider the decoding algorithm used most broadly for decoding HMMs – the Viterbi algorithm.

3. Viterbi algorithm

Viterbi algorithm for decoding HMMs (i.e. establishing the most probable sequence of hidden states, $\hat{S} = (\hat{s}_1, \hat{s}_2, \cdots, \hat{s}_t)$ from a sequence of observations *O*) relies on recursively identifying the most probable state at every step given the most likely hidden state at the previous step (this entire exposition is based on [10]).

The algorithm's operation relies on calculating two tables:

$$C_1(j,t) = \max_n \left(C_1(n,t-1) \cdot A_{nj} \cdot B_{jo_t}\right) \tag{1}$$

which describes, given previous observations, the highest probability of the hidden state transitioning to x_j and emitting observation o_t, and

$$C_2(j,t) = \arg\max_n \left(C_1(n,t-1) \cdot A_{nj}\right) \tag{2}$$

which describes the most likely (given previous observations) state transitioned from ($\hat{s}_{t-1}$) at step *t* given $\hat{s}_t = x_j$. The algorithm is initialised with

$$C_1(j,1) = \pi_j \cdot B_{jo_1}, C_2(j,1) = 0 \tag{3}$$

and run until the tables C_1 and C_2 are filled. Then, the hidden state corresponding to the greatest element in $C_1(j,T)$ is identified, so

$$\hat{s}_t = \arg\max_j \left(C_1(j,T)\right) \tag{4}$$

is the most likely ending state given all the observations, and the most likely previous hidden states are recursively chosen as $\hat{s}_{t-1} = C_2(\hat{s}_t, t)$.

Equations (1), (2), (4) show the conditions of reliable functioning of the algorithm (from (2)):

$$\forall t \in (1 \ldots T), j \in (1 \ldots N) \, \exists! \arg\max_n \left(C_1(n,t-1) \cdot A_{nj}\right) \tag{5}$$

as multiple maxima in any column of $C_1(n,t)$ make decoding ambiguous, because the state sequence cannot be chosen uniquely, and, from (4):

$$\forall t \in (1 \ldots T), j \in (1 \ldots N) \, \exists! \arg\max_j \left(C_1(j,t)\right), \tag{6}$$

which is a stronger version of the same requirement, but for $C_1(j,T)$ dictated by the necessity of being able to unambiguously establish the state most likely responsible for the final output. As we want to be able to decode sequences of arbitrary length, this requires us being able to terminate decoding at any step.

Let us consider why the trivial construction of *A* and *B* guarantees the failure of decoding. Substituting into equations (1), (2), (3) we obtain:

a) $C_1(j,1) = G(j) \cdot F(o_1)$,

b) $C_1(j,2) = \max_n \left(G(n) \cdot F(o_1) \cdot G(j) \cdot F(o_2)\right) = \max G(n) \cdot \left(F(o_1) \cdot G(j) \cdot F(o_2)\right)$,

so, by induction, every column of C_1 will be a scalar multiple of *G*. If there are multiple maxima in *G,* the algorithm fails at this stage by condition (5),

c) $C_2(j,2) = \underset{n}{\arg\max}\left(G(n) \cdot F(o_1) \cdot G(j)\right) = \arg\max\left(G(n)\right)$ – the outputs are always decoded as originating from the state with the highest probability.

Thus, we can describe another mode of failure, where regardless of observed outputs, the predicted state does not change. This is brought about by the outputs being independent of states and consecutive states being independent of each other, as is the case if the matrices are constructed in a trivial way.

In addition to the absolute proscriptions on the construction of the HMM described above, there are also several risk factors increasing the possibility of the derailment of the method: (a) multiple hidden states sharing the same steady-state probability (i.e. multiple identical elements of π), especially combined with other risk factors, (b) multiple identical large elements in columns of *A* and *B* increase the probability of multiple maxima in columns of $C_1(j,t)$, (c) multiple identical large elements in rows of *A* may lead to multiple maxima in $C_1(j,t)$ and failure as described by equation (6).

In addition, it must be borne in mind that the operation of the method relies on successive multiplications of probabilities, so the elements of C_1 will rapidly decrease, thus the possibility of two elements being considered equal due to computation precision when it is not the case increases with the length of the observation sequence. Therefore, for practical purposes, every instance of the word “identical” in this section can be replaced by “similar”.

Using these findings, the method of HMM construction was developed.

4. Constructing the HMM

First, one has to consider the message to be encoded. If each symbol has a distinct probability (for instance, if the symbols to be transmitted are letters in a natural language) the situation is simpler, as we only have to deal with repeated elements in columns of *A*. However, if the symbols have equal probabilities (for instance, if they represent bits), the problem of repeated elements in rows of *A* arises.

The number of hidden states can be increased by splitting one symbol into its combinations – for instance, introducing 01 and 00 instead of 0 when encoding digital message. This approach ensures the preservation of the capacity of the HMM to encode any signal, and can split the large maxima in *A* and increase symbols throughput. However, there is a limit on the relative number of hidden states and emitted states, which will be discussed later.

The solution of this problem is pre-seeding the message to be encoded with blank characters, with the probability of inserting the blank dependent on the current and next symbol – this can be represented by the matrix $D_{ij}, i, j = 1 \ldots N$, which represents the probability of inserting a blank signal if the current symbol is x_i and the next one is x_j. Then, the size of the original A is increased by one row and one column to accommodate the new blank symbol, and the new transition matrix A^* is constructed:

$$A^*_{ij} = \begin{cases} A_{ij}\left(1 - D_{ij}\right), \; j = 1 \ldots N, \; i = 1 \ldots N \\ G(i) \sum_j A_{ij} D_{ij}, \; j = N+1, i = 1 \ldots N \\ \left(\sum_i G(j) G(i) D_{ij} \right) / \left(\sum_i \left(G(i) \sum_j A_{ij} D_{ij} \right) \right), \; j = 1 \ldots N-1, i = N \end{cases} \tag{7}$$

with the remaining element $A^*_{N+1,N+1}$ equal to 0. The elements of D are selected to ensure acceptable distinctness of transitions of hidden states by Matlab® optimisation packages (under requirement of each row and column composed of distinct elements).

The emission matrix, just like the transition matrix, must be reconstructed from the trivial construction. Here, the distinctness of elements in columns is paramount in the current iteration of the method – the distinctiveness within rows would be important if we relaxed the assumption of perfect reception, or if we wanted to use the output from this stage as an input to another matrix.

The construction of the emission matrix also requires a software optimiser – for this paper, Matlab® optimisation package was used. The problem is set up thus: (a) $\sum_l b_{kl} = 1$ (emission at each hidden state), (b) $\sum_k \pi_k b_{kl} = F(l)$ (desired observation distribution), $\max_l \sum_k b_{kl} \cdot \log_M b_{kl} = \min$ (maximum distinctness of elements in columns), (c) each column composed of unique elements, and optimised using Matlab® optimisers. This problem involves $N \cdot M$ variables (the size of *B*), and (excluding pathological cases) $N + M - 1$ independent equations ($N - 1$ from condition a), M from condition b)). Therefore, for the problem to have a non-trivial solution, $NM > N + M - 1$, or $NM \geq N + M$. This equation describes the practical limit of splitting hidden states and addition of synonymous states. Naturally, the grater the degree of inequality, the more possibilities of optimisation exist, and the more the matrix elements can be tweaked. The initial distribution is selected as the left eigenvalue of A^*. It is assumed that we can perfectly distinguish between observed states.

5. Results

For the sake of comparison with a mimicry encoding method which relies on encoding messages in permutations of animal signals to give the output a Zipfian distribution [7], the hypothetical scenario of encoding 3 binary syllables (1, 01, 00 – 1.5 bits) in six vocal units (referred to as **1 2 3 4 5 6**) was explored. The frequency of units returned by the Zipfian method is presented in Figure 1 as compared to a purely Zipfian distribution.

Next, a HMM was constructed to mimic the same scenario: 3 binary syllables and the filler syllable (last row and column of A^*), and an exactly Zipfian distribution of six output units (frequencies of 0.408, 0.204, 0.136, 0.102, 0.0816, 0.068 – squares in Figure 1). The model definition was then selected:

$$D = \begin{bmatrix} 0.9963 & 0.1814 & 0.3918 \\ 0.0442 & 0.9970 & 0.1099 \\ 0.0112 & 0.0868 & 0.9939 \end{bmatrix} \quad A^* = \begin{bmatrix} 0.0018 & 0.2046 & 0.1521 & 0.6415 \\ 0.4779 & 0.0008 & 0.2225 & 0.2988 \\ 0.4944 & 0.2283 & 0.0015 & 0.2758 \\ 0.5513 & 0.1947 & 0.2540 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} 0.0004 & 0.5406 & 0.0970 & 0.2805 & 0.0807 \\ 0.5497 & 0.1061 & 0.0040 & 0.0145 & 0.0015 \\ 0.0027 & 0.0018 & 0.5952 & 0.0172 & 0.2316 \\ 0.9893 & 0.0034 & 0.0021 & 0.0026 & 0.0020 \end{bmatrix} \quad \pi = \begin{bmatrix} 0.3414 \\ 0.1707 \\ 0.1707 \\ 0.3171 \end{bmatrix} \tag{8}$$

Afterwards, tests of reliability of encoding and decoding were undertaken – observation sequences between 2 and 2000 characters long were generated and decoded, with their error rate (frequency of errors in state decoding) averaged over 20 runs for each sequence length. The results are presented in Figure 2. In addition, a test of decoding a 2000-observation sequence split into blocks of length 10 is presented in the same figure.

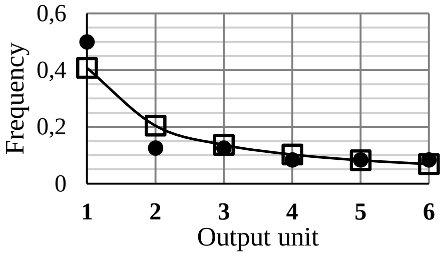

Figure 1. Zipfian method output unit frequency (black circles) and Zipfian distribution for six elements (black squares)

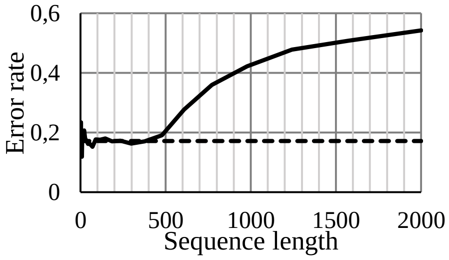

Figure 2. Decoding error rate for HMM for varying sequence length (solid line) and error rate for decoding a long sequence in blocks of 10 units (dashed line)

6. Discussion

Based upon the results (Figure 2), one notices that the HMM method is inherently error-prone – we can see two regimes of its operation: for a range of decoded observation sequence lengths, the error rate is stable and relatively low, transitioning to a much higher value for longer sequences.

The fact that any error can be explained thus: the Viterbi algorithm gives the most probable sequence of hidden states, but sometimes the improbable transition will happen, throwing off the decoding process. This risk is not present in the Zipfian mimicry method, as each vocal unit permutation can only correspond to one message block.

The jump in error rate is due to one of the problems described in section 0 – namely, the successive multiplication of probabilities leading to the method not being able to establish unique maxima at consecutive steps of calculations, eventually decoding every observation as emitted from the most likely hidden state. However, this behaviour of the method does not preclude it from being used to form and decode long chains of units, such as those present in dolphin bursts [9], since the observation sequence can always be decoded in short blocks, which guarantee operating in the nominal, low-error rate value. This is because the initial state distribution is defined as the steady-state distribution of states, which eliminates any dependence of the decoded message on the chosen exact starting point. This also means that a portion of the message will always be accurately decoded even if not all vocal units are received.

Both errors can be decreased by constructing the HMM in a way that maximises the distinctness of state transitions – at the moment, this is done intuitively, a more in-depth mathematical treatment of the problem would be highly beneficial. However, there is a practical limit to this improvement, as eventually, for long sequences, computational precision will make decoding impossible – even a relatively large probability of 0.5 taken to the 500th power (a qualitative analogue of decoding the 500th observation) is of the order of $3 \cdot 10^{-151}$ – well below the computational precision of Matlab® (32 digits by default).

The quality of mimicry described by both methods is very different. The Zipfian method can only approximate one (Zipfian) distribution, always trading the number of redundant vocal units for the closeness of the distribution to the desired one. In addition, in the Zipfian method, the individual meaning-carrying permutations of units must be separated either by a pause or by a non-coding vocal unit, thus either introducing a hard limit on sequence length or radically reducing the information throughput (the separator unit is guaranteed to be the one with the greatest frequency, as it appears in every permutation). In contrast, the HMM method can operate on arbitrary, even non-Zipfian vocal unit distributions, and can reliably generate a vocal sequence of arbitrary length. While it is true that a some symbols will be decoded incorrectly, this problem can be alleviated by using error-correcting codes (such as convolutional codes) to pre-code the message.

The pre-seeding of the message with blanks, while necessary for improving the robustness of encoding and decoding, decreases the bitrate (in the simulated HMM – to 1.02 bits/vocal unit from 1.5 in the Zipfian method). It would be beneficial to investigate the exact trade-off between bitrate and distinctness of the transition matrix further and introduce a measure of weighing bitrate against distinctness, as was done in [7]. It must also be borne in mind that the scenario of encoding bits was chosen in this paper precisely because it tends to be challenging due to multiple hidden states appearing with the same probability – the method is general enough to encode other signals, e.g. letters of alphabet or Morse code, which are far less challenging due to the each signal, having a different probability of occurring in most cases.

The question of whether transition and (especially) emission matrices meeting the criteria of uniqueness of elements can be constructed for a general distribution of message symbols and output vocal units is still unanswered, and would benefit from further mathematical study – to be frank, we know how to make the method not fail immediately and intuit how to make it robust, but there is no mathematical clarity as to what optimisation strategy would be the best or even if optimisation will always succeed.

Finally, the HMM approach to mimicry looks like a promising avenue of research, and it appears very likely that further stacking of HMMs will enable one to simulate not only desired vocal unit distributions, but also entire conditional structures of animal species' communication. Also, the possibility of combining the Zipfian method's approach of introducing synonyms will be investigated, as will the relaxation of the assumption of perfect distinction between vocal observations/emissions.

7. Conclusions

The HMM-based method of mimicry for covert communication can reliably encode and decode messages while mimicking the desired vocal unit frequency distribution.

The HMM-based method of mimicry can operate reliably even when decoding long or incomplete chains of observations.

References

1. S. Liu, G. Qiao, A. Ismail, *Covert underwater acoustic communication using dolphin sounds*, J Acoust Soc Am, **133**(4) (2013) EL300 – EL306.
2. A. Kershenbaum et al., *Acoustic sequences in non-human animals: a tutorial review and prospectus*, Biol Rev, **91**(1) (2016) 13 – 52.
3. B. McCowan, S. F. Hanser, L. R. Doyle, *Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires*, Anim Behav, **57** (1999) 409 – 419.
4. R. Ferrer-i-Cancho, B. McCowan, *A law of word meaning in dolphin whistle types*, Entropy, **11** (2009) 688 – 701.
5. L. R. Doyle. *Animal Communications, Information Theory, and the Search for Extraterrestrial Intelligence (SETI)*. [Online]. https://www.seti.org/seti-institute/animal-communication-information-theory-and-seti.
6. J. Kanwal, et al., *Zipf's law of abbreviation and the principle of least effort language users optimise a miniature lexicon for efficient communication*, Cognition, **165** (2017) 45 – 52.
7. K. W. Zuber, K. J. Opieliński, *Animal mimicry in covert underwater communication application of syntax generation and simulated genome method*, in 2018 Joint Conference - Acoustics, Ustka, 2018, 1 – 5.
8. K. Katahira et al., *Complex sequencing rules of birdsong can be explained by simple hidden Markov processes*, PLoS ONE, **6**(9) (2011) 1 – 9.
9. A. R. Luis, M. N. Couchinho, M. E. dos Santos, *A quantitative analysis of pulsed signals emitted by wild bottlenose dolphins*, PLOS one, **11**(7) (2016) 1 – 11.
10. L. R. Rabiner, *A tutorial on hidden Markov models and selected applications in speech recognition*, Proc. of the IEEE, **77**(2) (1989) 257 – 286.