Supplementary material for:

Imitation of novel conspecific and human speech sounds in the killer whale *(Orcinus orca)*

José Z. Abramson, Mª Victoria Hernández-Lloreda, Lino García, Fernando Colmenares, Francisco Aboitiz & Josep Call

Supplementary Material 1: Details of Methods for training Phase 2

In the training phase three familiar vocalizations were used. One of them ('Song') had been trained, was regularly requested for all subjects to perform during the aquarium shows with a specific hand sign for it and was the only sound used in our previous action imitation study. The other two ('Birdy' and 'Blow') were part of their natural repertoire, but had been trained to perform on command (conditioned to a specific hand signal) just for the model. In this second phase, the subject was positively rewarded with fish and with tactile and voice reinforcement signals whenever she yielded a correct response to the model's familiar sounds. She received no reinforcement following errors. Familiar sounds were judged in real time by two observers, (Wikie's trainer and one experimenter) as the sounds were considerably different and readily audible, thereby allowing the experimenter and the trainer to correctly distinguish them by listening. Reinforcement of the model was not contingent upon the response of the subject. In the first sessions we used the sound 'Song' and introduced for the first time the familiar sound 'Birdy', a sound that on occasions the subject naturally produced after being fed (and, therefore, we speculated it could be associated to a positive emotional state). The subject didn't correctly make the transfer after three sessions, as she only produced the sound 'Song' when signalled to copy 'Birdy'. As 'Birdy' was not a trained sound, we decided to drop 'Birdy' and introduce another familiar sound, 'Blow', which differed from 'Birdy' in that the former was already trained (associated to a specific signal), and was regularly signalled in veterinary procedures. We introduced the sound 'Blow' first with the copy signal followed by its specific hand signal for

one session and then in a second session we presented ‘Blow’ with the copy signal alone. The subject successfully copied the ‘Blow’ sound and the ‘Song’ sound just with the copy signal alone in this second session. Finally, in another session we introduced again the ‘Birdy’ sound, adding this third familiar sound to the others, and the subject correctly made the transfer for all these three familiar sounds. Her matching for ‘Birdy’ was accurate from the sixth session.

For the human-made sounds condition we ran two previous transfer sessions to the human model where the human model performed three already tested sounds, the two familiar sounds ‘Song’ and ‘Blow’ and the novel killer whale sound ‘Strong Raspberry’ (that was now familiar to the subject as it had already been tested and found to be performed accurately).

The model’s training with novel sounds was done in an isolated pool, away from the experimental subject, with standard conditioning procedures by capturing novel spontaneous utterances from the subject and then shaping these sounds and associating them to a new specific hand signal.

General Testing Schedule

During experimental sessions, subjects were not food deprived, and testing was interrupted if they refused to participate. Experimental sessions were conducted typically between 10:00 a.m. and 12:00 p.m. and between 15:30 p.m. and 17:00 p.m. Each session consisted of 6 -15 testing trials and 10-25 familiar sounds control trials, given a total of 15-40 (up to 50 trials, depending on the motivation of the subject to participate and the ‘trainers schedule’), lasting approximately 20–30 min altogether. There were one to three sessions per day, four days a week. Some sessions were finished earlier if subjects were distracted or disinclined to participate.

Sound categories.

The sounds categories performed by models were: (a) familiar killer whale sounds emitted by a conspecific live (phase 1); (b) familiar killer whale sounds emitted by a speaker (phases 1 and 2) (b) novel killer whale sounds emitted by a conspecific live model (phase 3); (c) novel killer whale sounds emitted by a speaker (phase 3); and (d) human sounds (Phase 3). Two novel sounds uttered by the conspecific model ('Wolf' and 'Elephant') were only presented via the speaker and not in the conspecific live condition because they were not reliably produced by the live model. Table S1 gives the complete list of sounds and their description

When defining novelty it's important to note that it cannot be reduced to an all-or-nothing matter as there will always be some degree of similarity to what the imitator has done before (since an individual will probably have produced, at one time or another, all of the muscular movements of which it is physically capable, which makes novelty an issue of recombination of actions) (Whiten and Custance 1996). However, to asses that our unfamiliar sounds (conspecific and humans) were as different as possible of what they produced before we compared them with 278 sound samples facilitated kindly to us by Hodgins and colleagues. This samples were extracted from her master thesis sound recording baseline in this same killer whale group housed at Marineland, where she identified eleven distinct discrete call types, and we didn't find any sound similar to a conspecific or human novel sound. Complementary, we also record in air 28 hours of the killer whales during their free time before running the experiment to see if the subject (or any other killer whale in the group) uttered sounds similar to the novel sounds. In this 28 hours of recordings only four events/instances of above the surface vocalizations. This spontaneous vocalizations where identified as vocalization Nº 7 recorded by Hodgins 2005, and identified as a "stress" or as a "discomfort" vocalization from the whales, therefore we didn't

recorded any event previous to the experiment in which the whales uttered a vocalization that resembles to the novel vocalizations we used in the test.

Table S1. Sounds list and description

	Description
FAMILIAR SOUNDS	
Song (SO)	A strong moan, kind of whine tonal sound
Birdy (BI)	A tiny volume modulated high frequency sound with tonal variation made of small peeps similar to a bird call or to the song of a cricket or a cicada
Blow (BL)	The natural atonal low frequency sound produced during breathing process by expelling and then inhaling the air from the animal's blowhole
NOVEL SOUNDS	
Conspecific	
Strong Raspberry (SR)	A strong noisy modulated and low frequency burst atonal sound
Creaking Door (CD)	A squeak type of modulated atonal low frequency sound similar to a rusty door lock or to the creak of a wooden rattle
Breathy Raspberry (BR)	A soft airy burst-pulse atonal sound
Wolf (WO)	A two ascending and then descending siren like tonal sound
Elephant (EL)	A strong volume modulated tonal sound made of whines and chirps similar to an elephant call
Human	
Ah Ah (AA)	A strong human laugh
Hello (HE)	Human words
Bye Bye (BB)	Human words
Amy (AM)	The name of the model trainer of the present study
One–Two (OT)	Human words
One-Two–Three (OTT)	Human words

Complete list of features selected for DTW: 1) *Spectral Pitch Contour ACF* (Autocorrelation Function of the Magnitude Spectrum), that shows the evolution of the fundamental frequency over time; 2) *Time Energy Evolution*: it allows comparing the evolution of the energy pattern in time between the model's and the subject's acoustic signal (temporal regularity and rhythm)**;** 3) *Pitch Class profile;* a histogram-like 12-dimensional vector (corresponding to the 12 notes of the diatonic musical scale) with each dimension representing both the number of occurrences of the specific pitch class in a time frame and its energy or velocity throughout the analysis block [41]. *4*) *Spectral Mel Frequency Cepstral Coefficients*, compact description of the shape of the spectral envelop of audio signal; 5) *Spectral Kurtosis* ("tailedness" of the probability distribution of a real-valued random variable), and 6) *Pitch Time AMDF,* computes the lag of the average magnitude difference function.

Table S2. Lowest DTW scores among the random sample of five copies of each vocalization type utilized for the DTW distance similarity index scale

	DTW	Similarity index (DTW/1000000)
FAMILIAR SOUNDS		
SO	105026	0,105026
BL	16542	0,016542
BI	117362	0,117362
NOVEL SOUNDS		
Conspecific		
BR	21897	0,021897
SR	28758	0,028758
EL	39776	0,039776
WO	41797	0,041797
CD	61251	0,061251
Human		
AM	33740	0,03374
HE	55427	0,055427
AA	81038	0,081038
OT	146902	0,146902
OTT	155034	0,155034
BB	199025	0,199025
Anchors (Benchmarks) scores		
Dissimilarities anchors: Incorrect random pair copies		
AM - OTT	940378 *	0,940378
BB - BR	302287	0,302287
CD - HE	249581*	0,249581
SR - HE	260794	0,260794
Similarities anchors: High quality copies		
HE with itself	0 *	
HE-HE (human subject)	24275 *	0,024275

Supplementary Material 2: Data figures.

ESM Figures in .eps available at 10.6084/m9.figshare.5446504

- Figure S1 Here we show one DTW example for the remaining familiar sound and for each of all the others novel sounds tested utilized for the DTW distance similarity index scale (conspecific and human).
- Figure S2-S4 Here we show one example of the main features selected for the DTW acoustic analysis for each novel sound tested. Audio samples both of each demonstrator's

(model) novel sound and subject's (imitator) copy selected for these acoustic analyses are also available as Supplementary Audio Files.

- Figure S2 (2.1 to 2.3). **Killer whale sound (*atonal*)**: *Wave form and spectrogram* of the model (a1) and imitated vocalization (a2); *Time energy distribution* of the model (b1) and the imitated vocalization (b2); *Chromagram* of the model and the imitated vocalization (c)
- Figure S3 (3.1 and 3.2). **Killer whale sound (*tonal*)**: *Wave form and spectrogram* of the model (a1) and imitated vocalization (a2); *Time energy distribution* of the model (b1) and the imitated vocalization (b2); *Fundamental frequency contour distribution* of the model (c1) and the imitated vocalization (c2); *Chromagram* of the model and the imitated vocalization
- Figure S4 (4.1 to 4.5). **Human sounds (*tonal*):** *Wave form and spectrogram* of the model (a1) and imitated vocalization (a2); *Time energy distribution* of the model (b1) and the imitated vocalization (b2); *Fundamental frequency contour distribution* of the model (c1) and the imitated vocalization (c2); *Chromagram* of the model and the imitated vocalization

Figure S1 Dynamic Time Warping for all the others sounds tested.

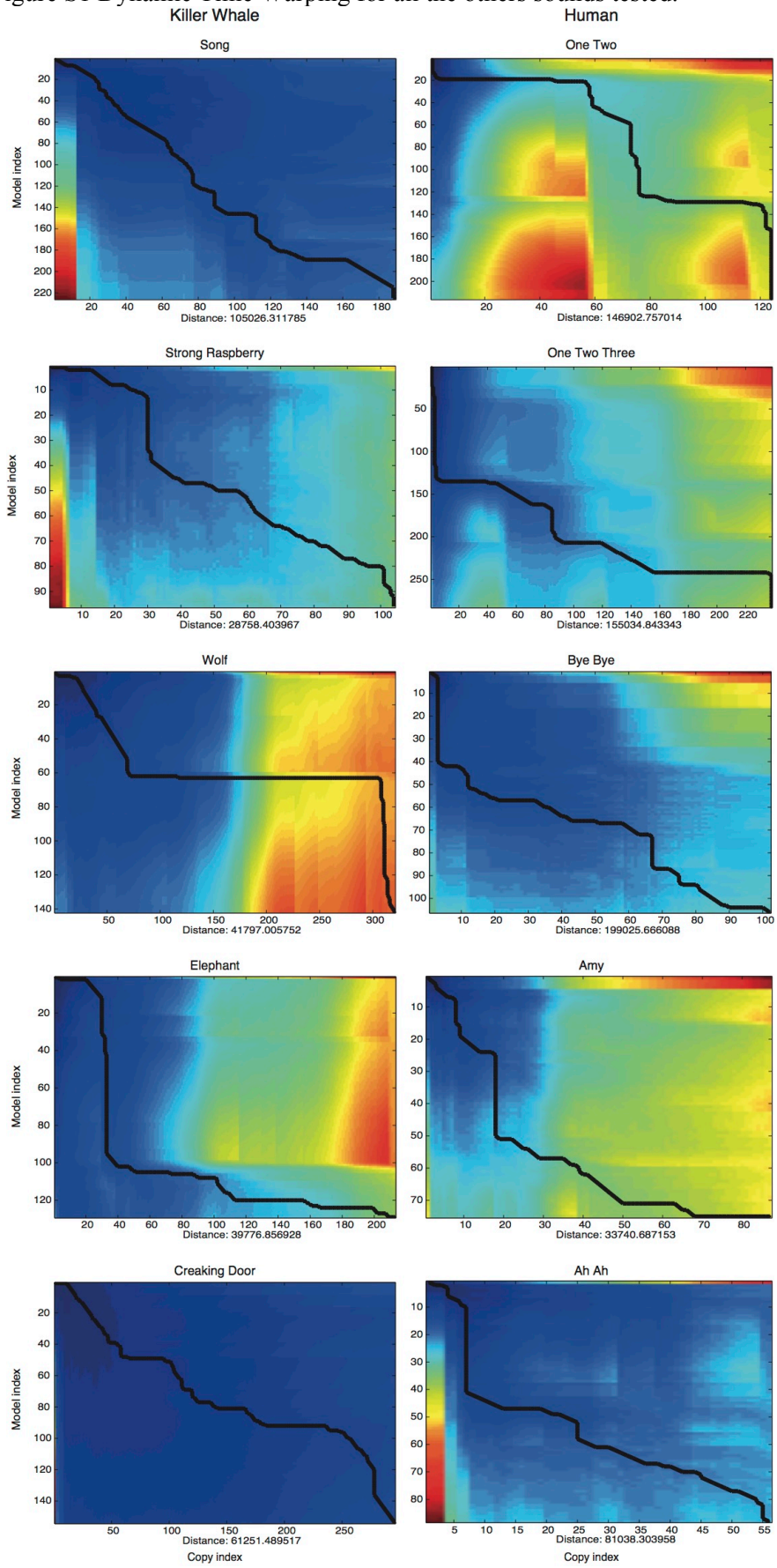


Figure S2.1

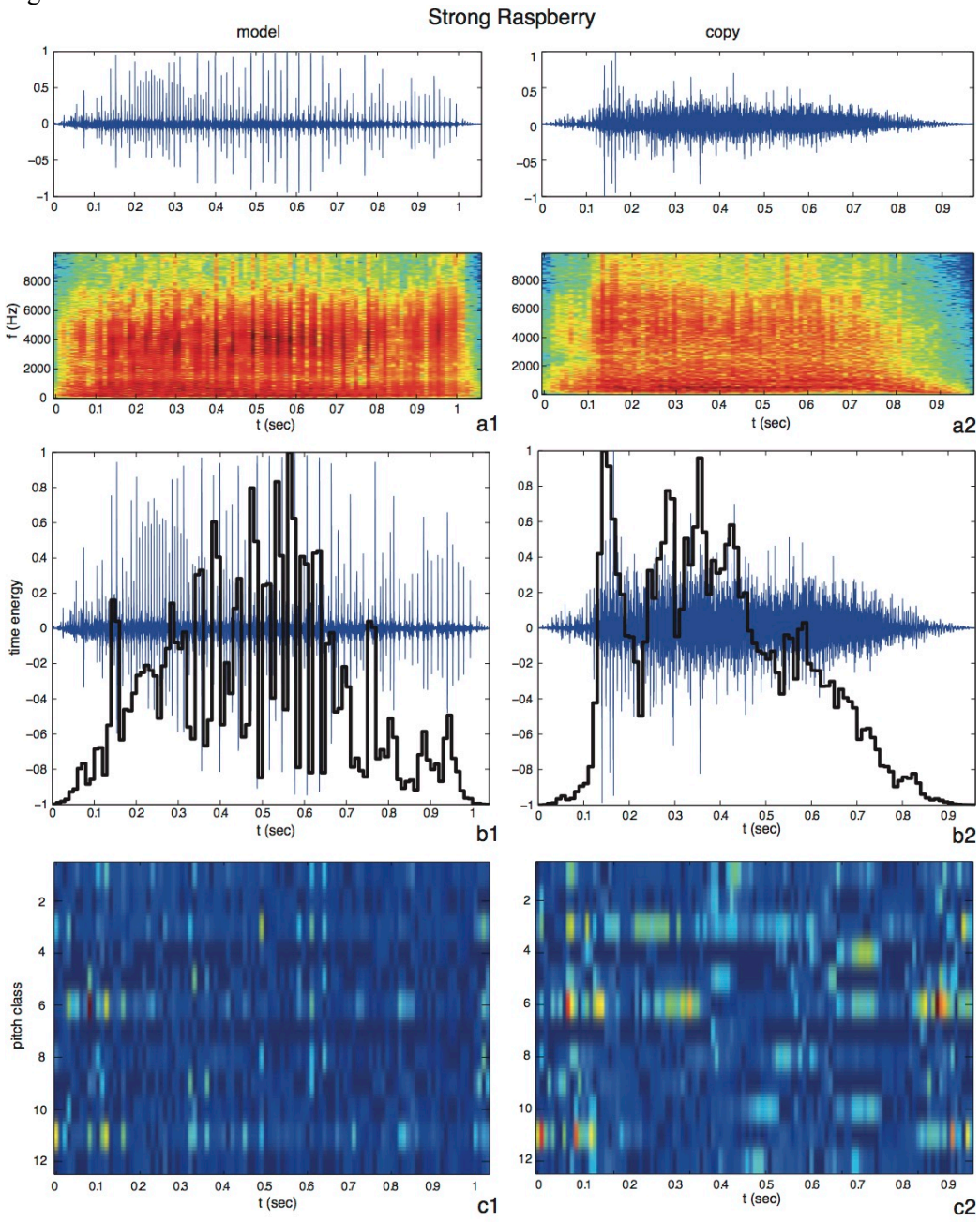

Strong Raspberry
model
copy
f (Hz)
t (sec)
a1
a2
time energy
b1
b2
pitch class
c1
c2

Figure S2.2

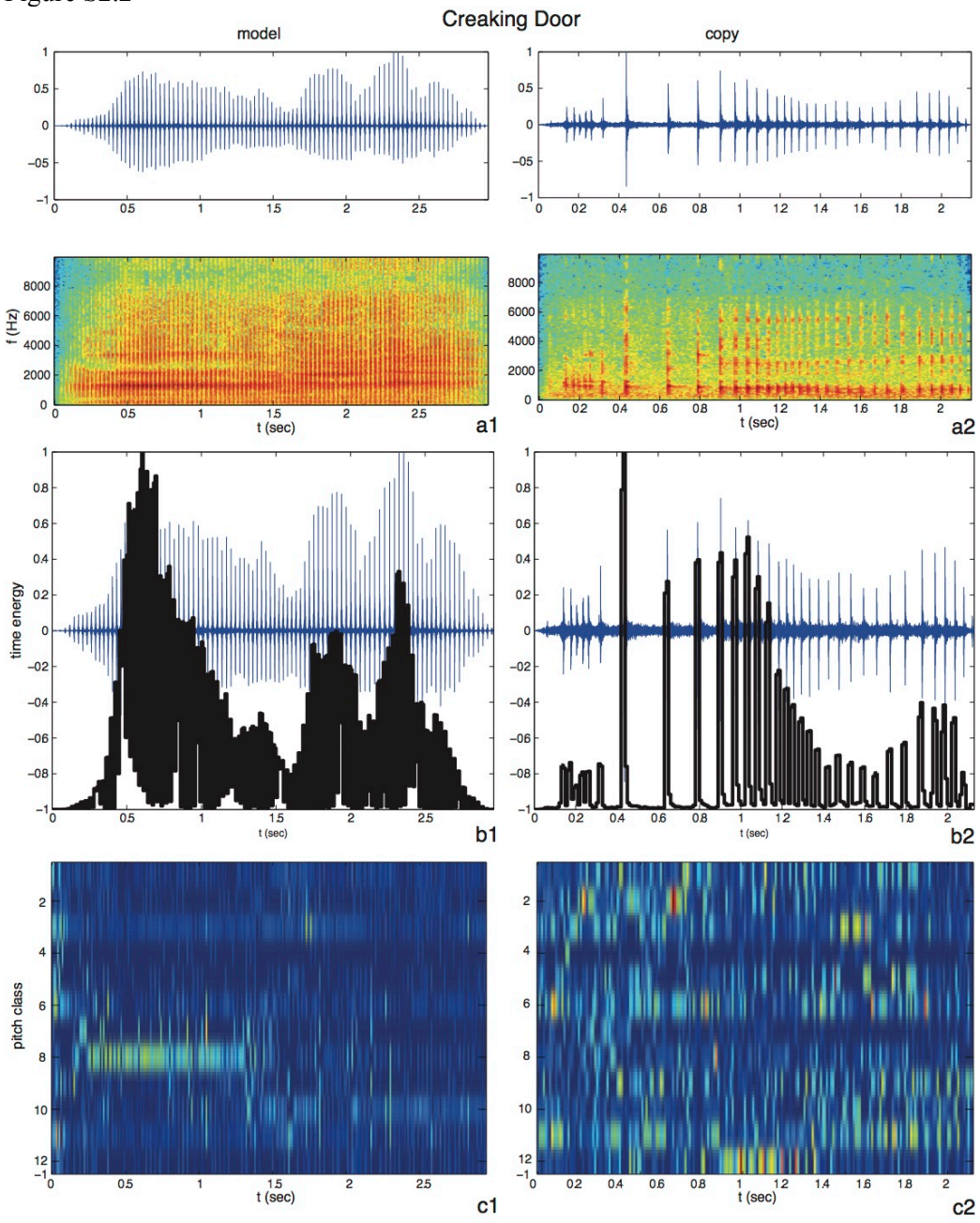

Creaking Door
model
copy
f (Hz)
t (sec)
a1
a2
time energy
b1
b2
pitch class
c1
c2

Figure S2.3

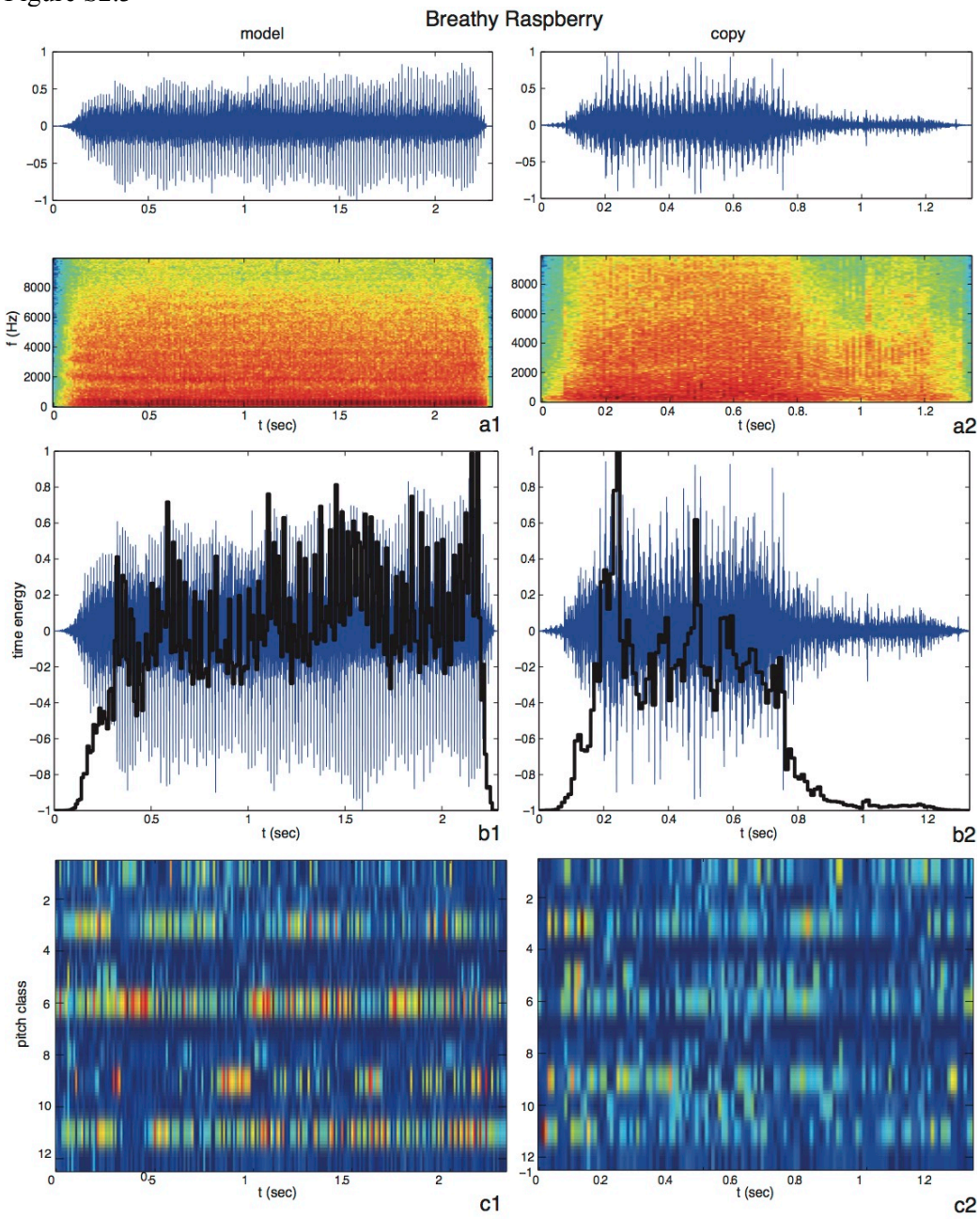

Breathy Raspberry
model
copy
f (Hz)
t (sec)
a1
a2
time energy
b1
b2
pitch class
c1
c2

Figure S3.1

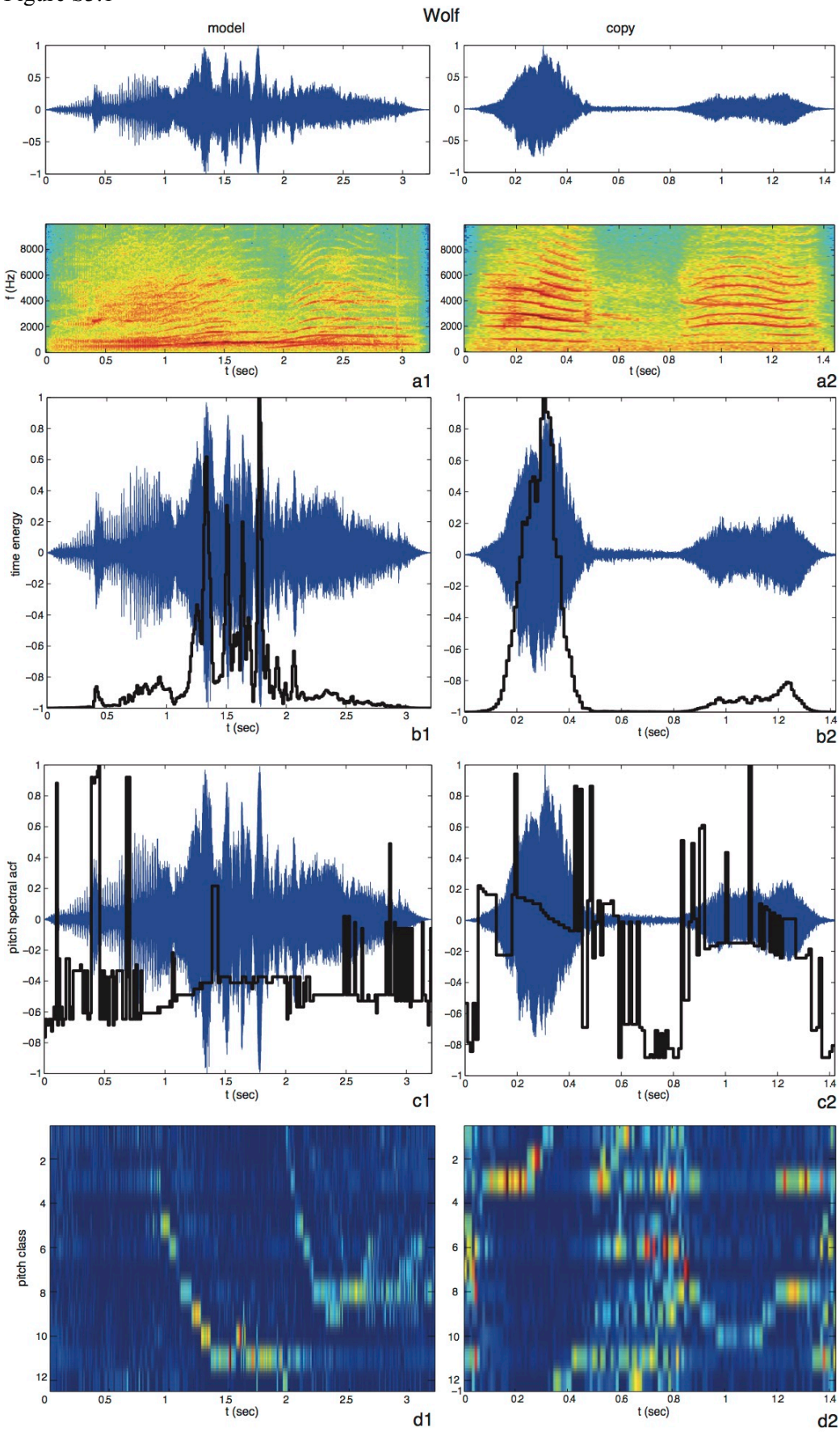

Wolf
model
copy
f (Hz)
t (sec)
a1
a2
time energy
b1
b2
pitch spectral acf
c1
c2
pitch class
d1
d2

Figure S3.2

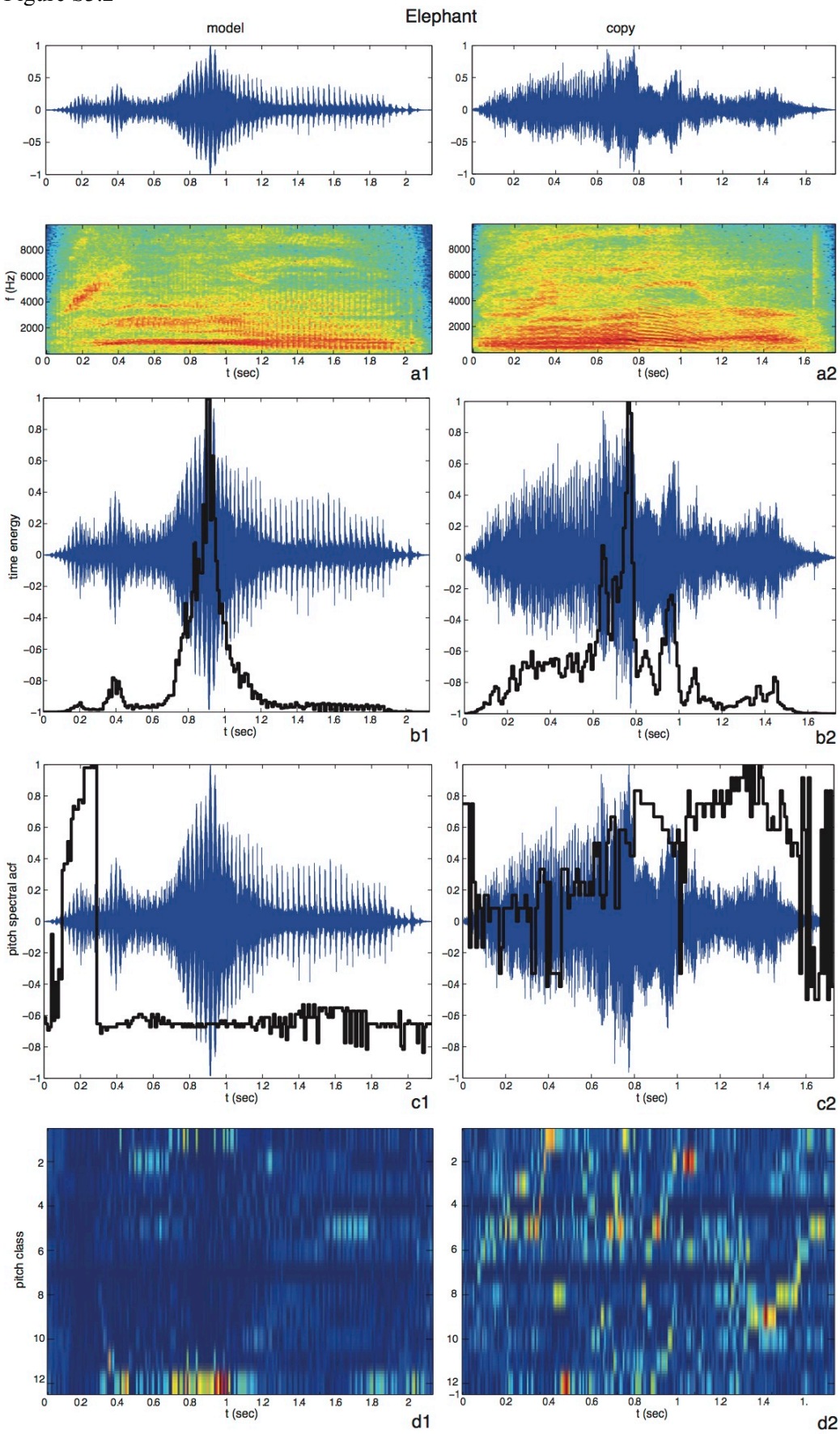

Elephant
model
copy
f (Hz)
t (sec)
a1
a2
time energy
b1
b2
pitch spectral acf
c1
c2
pitch class
d1
d2

Figure S4.1

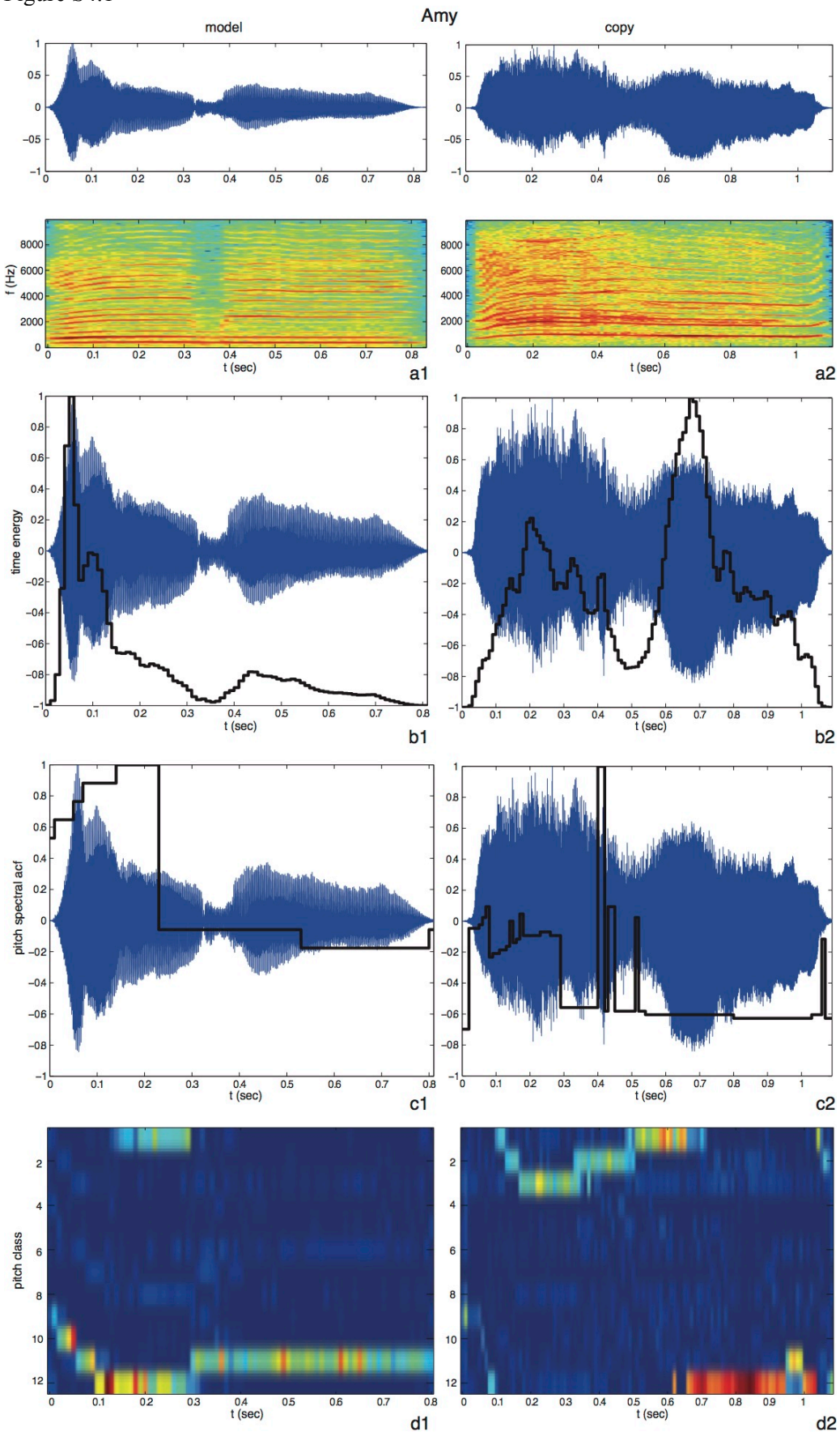

Amy
model
copy
f (Hz)
t (sec)
a1
a2
time energy
b1
b2
pitch spectral acf
c1
c2
pitch class
d1
d2

Figure S4.2

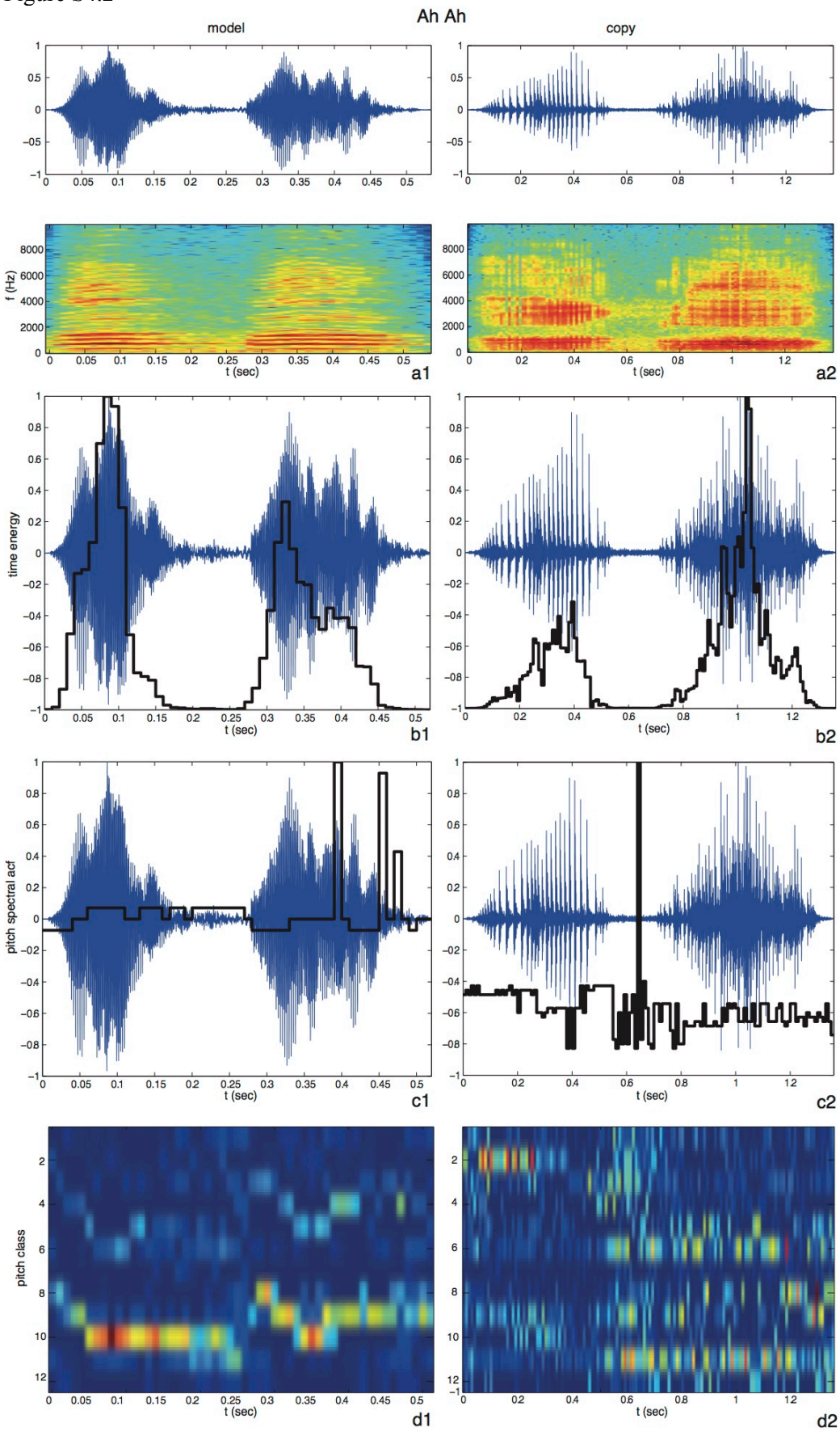

Ah Ah
model
copy
f (Hz)
t (sec)
a1
a2
time energy
b1
b2
pitch spectral acf
c1
c2
pitch class
d1
d2

Figure S4.3

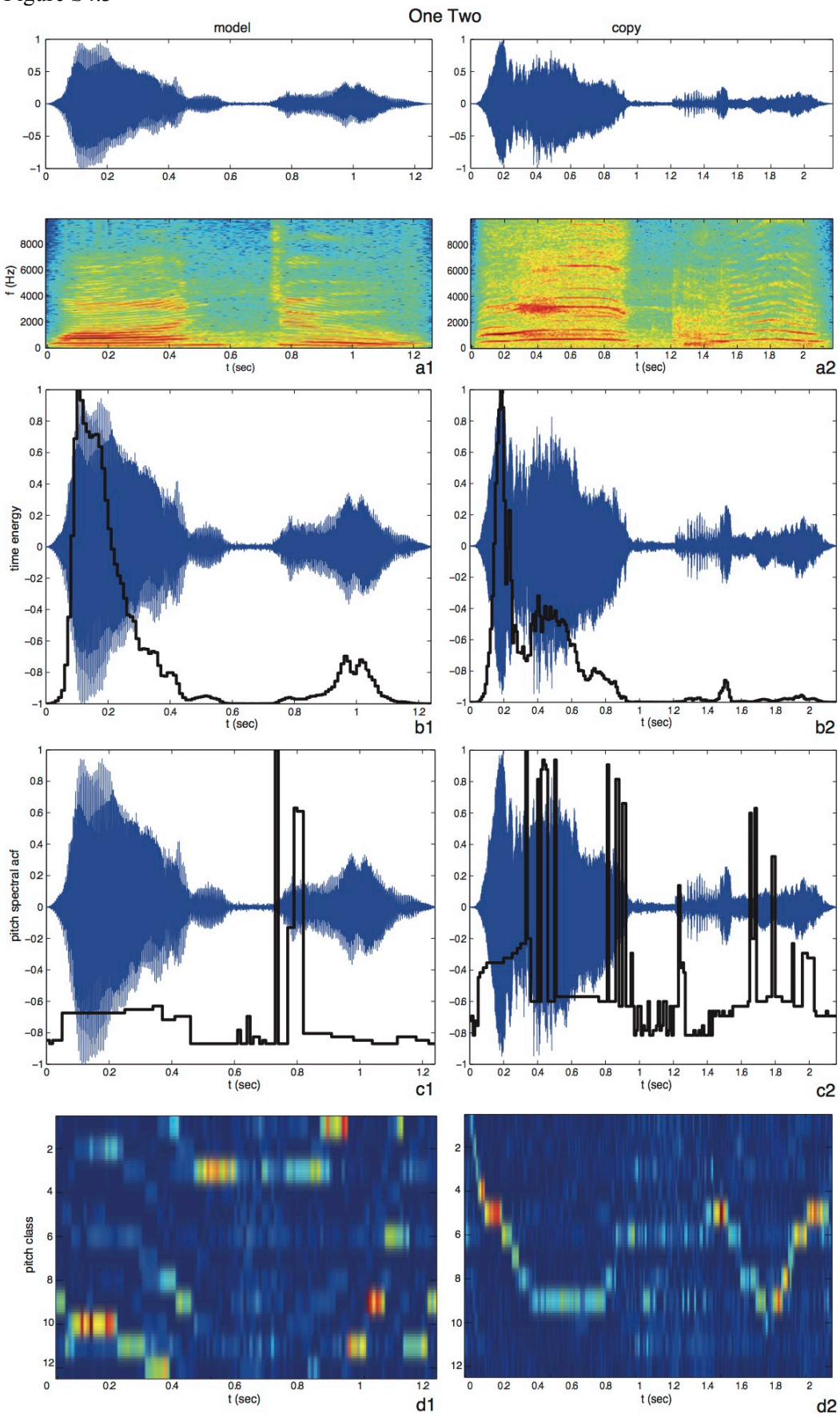

One Two
model
copy
f (Hz)
t (sec)
a1
a2
time energy
b1
b2
pitch spectral acf
c1
c2
pitch class
d1
d2

Figure S4.4

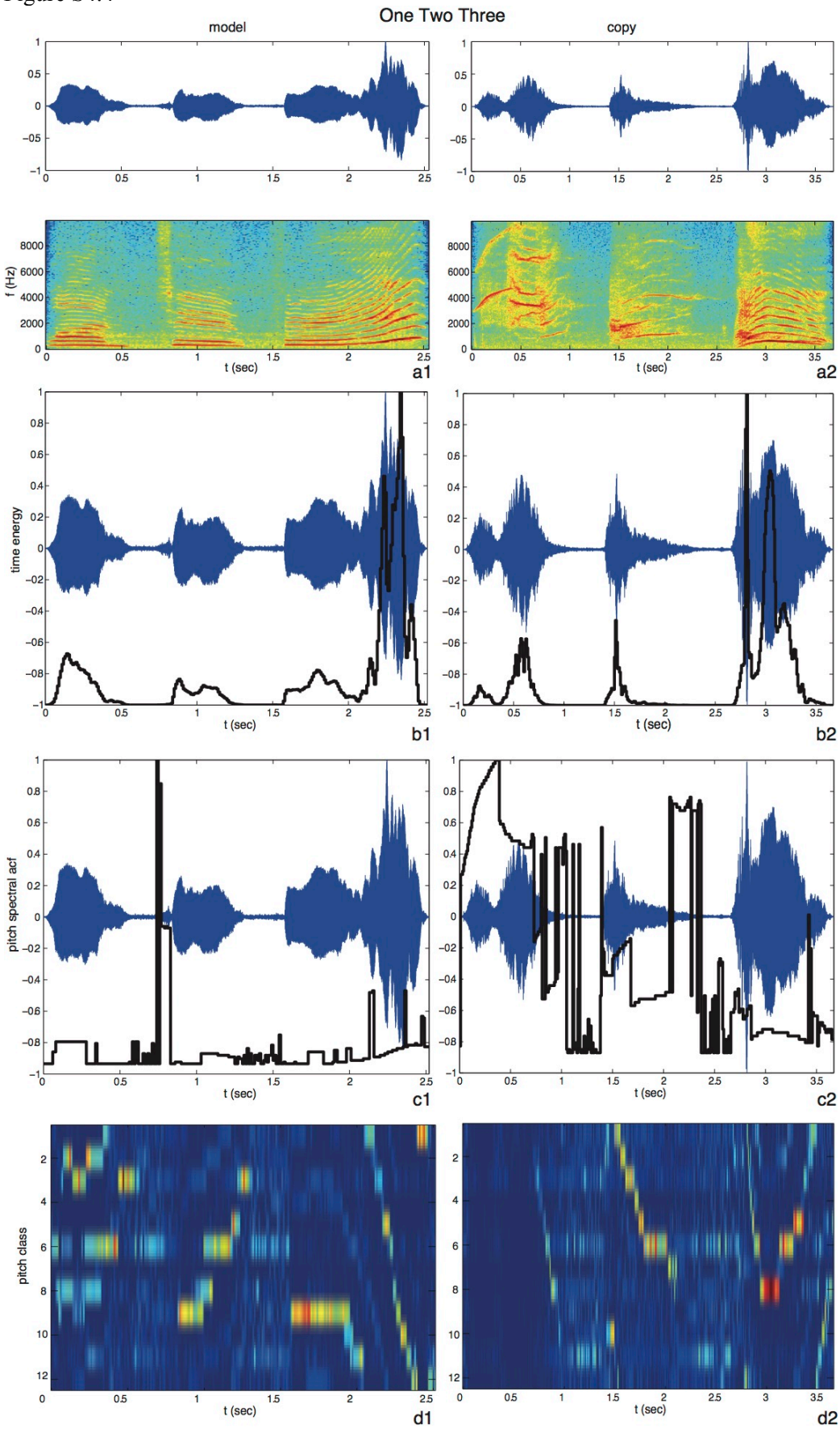

One Two Three
model
copy
f (Hz)
t (sec)
a1
a2
time energy
b1
b2
pitch spectral acf
c1
c2
pitch class
d1
d2

Figure S4.5

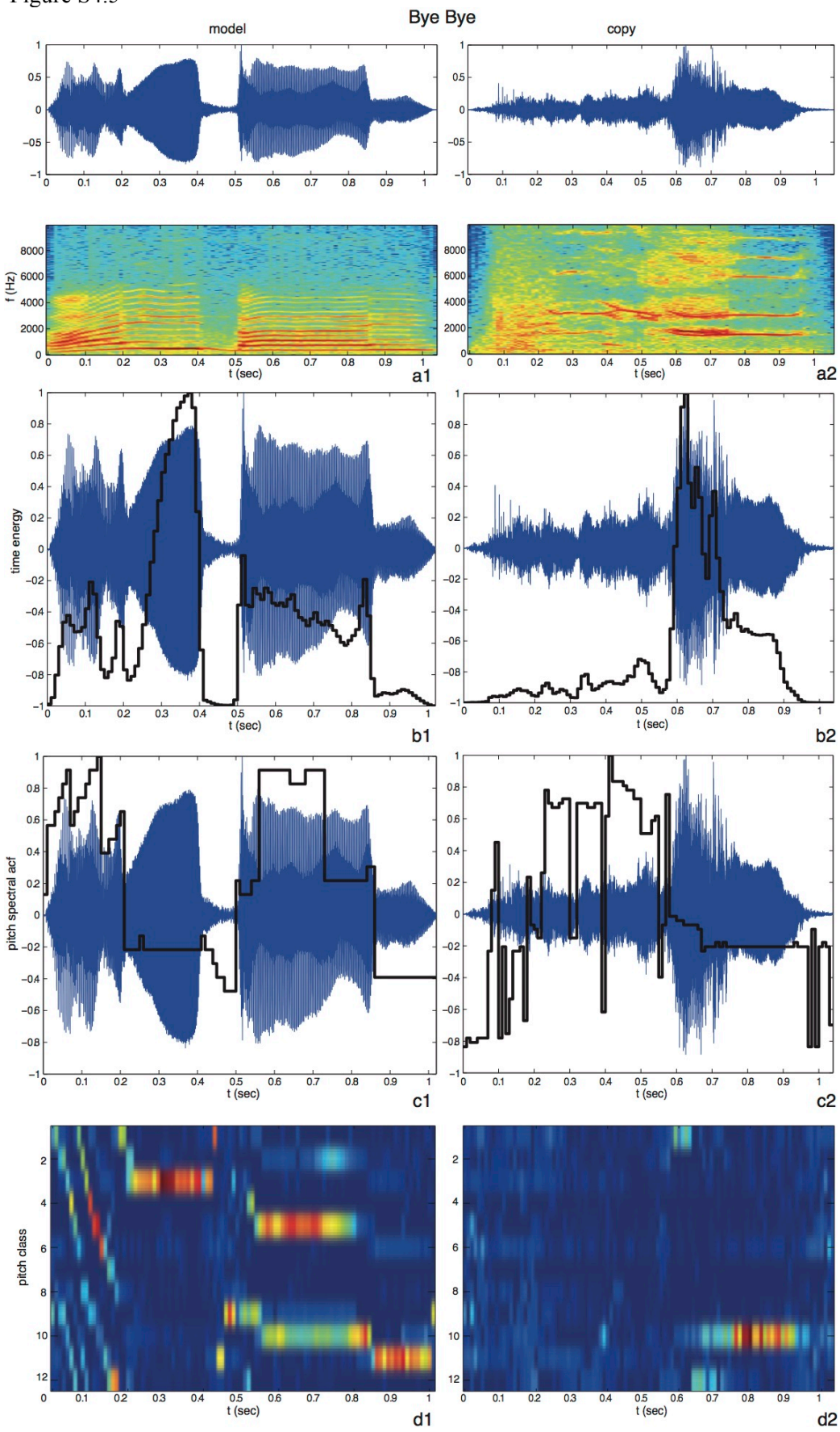

Bye Bye
model
copy
f (Hz)
t (sec)
a1
a2
time energy
b1
b2
pitch spectral acf
c1
c2
pitch class
d1
d2

Figure S4.6

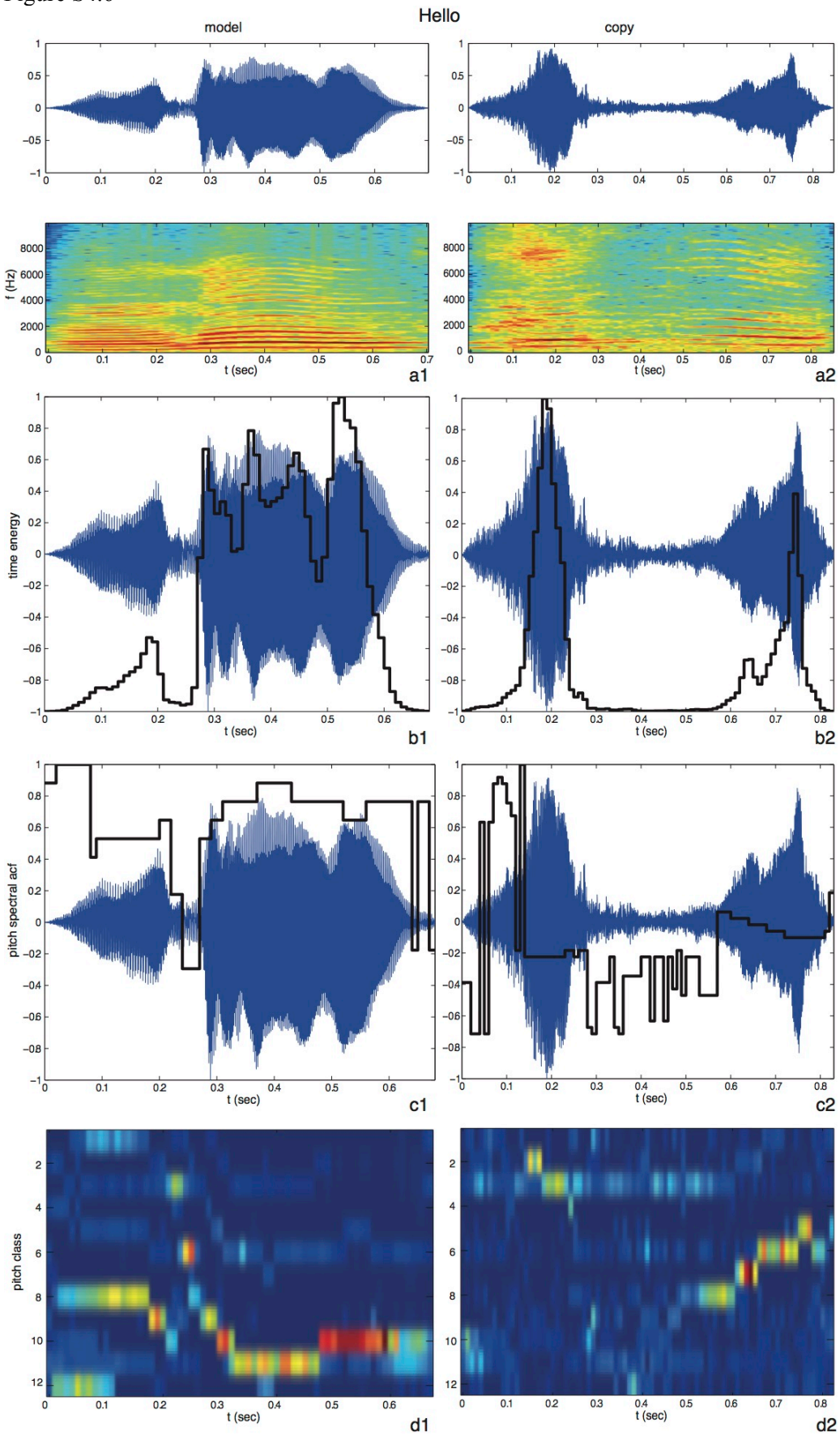
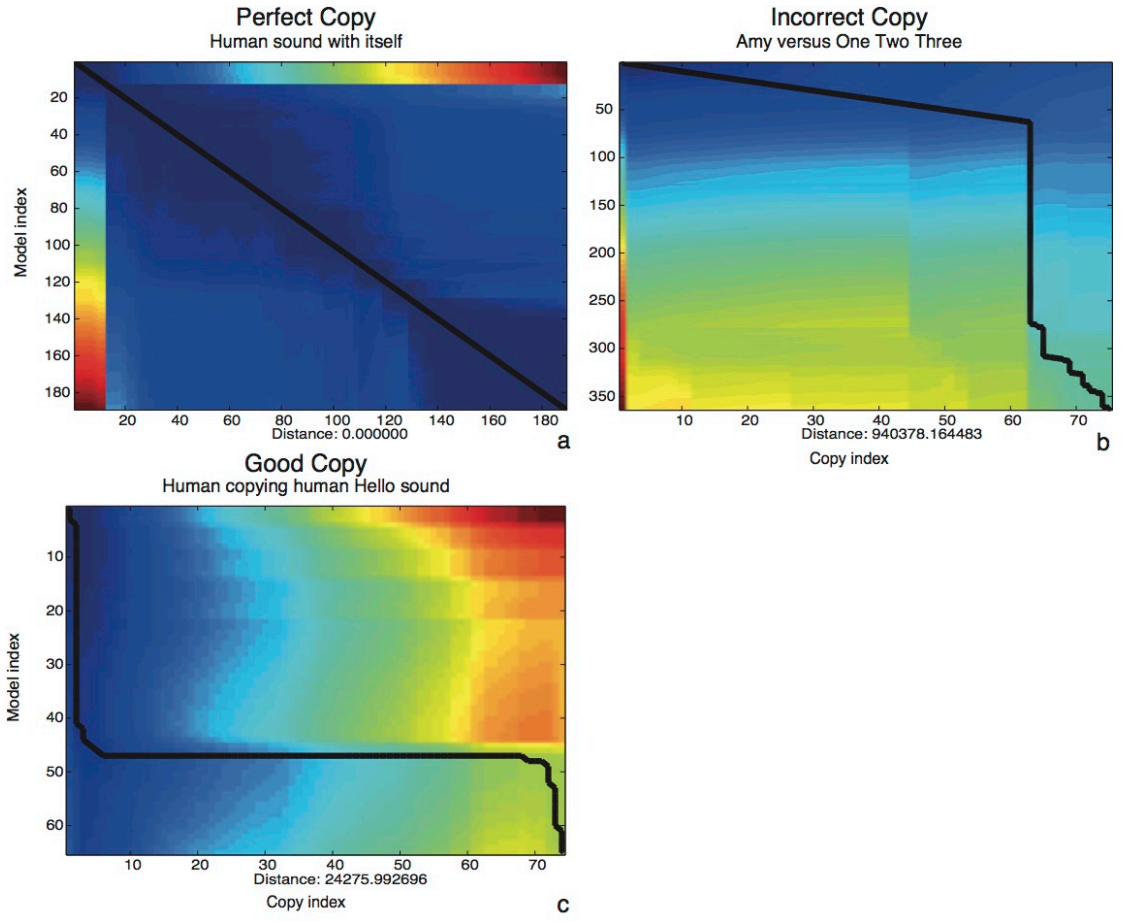

Hello
model
copy
f (Hz)
t (sec)
a1
a2
time energy
b1
b2
pitch spectral acf
c1
c2
pitch class
d1
d2

Figure S5 Dynamic Time Warping for the sounds used as anchors for the similarity index scale



[1]**Figure S5. Dynamic Time Warping of the benchmarks or "anchors" used to rescale the index into interval 0, 1.** (a); DTW of a perfect copy (i.e. a vocalization with itself) which corresponds to 0 in the scale (b); DTW of the maximum value found in demonstration sounds paired with copies that corresponded to other different model which in this case corresponded to the sound 'Amy' (tonal) of the demonstrated and the sound "One Two Three" (tonal) of the copy, which corresponds to the maximum dissimilarity value found from where the closest round value above corresponds to 1 in the scale. (c) DTW of a "good copy" or "high quality match" score (i.e. a human copying another human known word), which in this case corresponded to the sound "Hello" produced by the trainer and the experimenter copy of the same sound.

Supplementary Material 3: Audio Files Description

Audio file samples of each familiar and novel sound tested are provided in Supplementary Data. (Sampling frequency 48.000 Hz). For killer whale sounds, each example (copy trials) in the audio files is separated by a pure sinusoidal tone of 1 second and frequency of 440 Hz. For killer whale familiar sounds, we present one audio file that contains first the demonstrator's (model) sound and then one example (copy trial) of each of the 3 different familiar sounds from the training phase. For killer whale novel sounds we present five audio files that corresponds to each of the five novel sounds used in the testing phase. Each file contains first the demonstrator's (model) sound and then a random sample of 5 different examples (copy trials), where the first example is the copy chosen for the analysis in the manuscript and the ESM material. Finally for human novel sounds, we present six audio files that correspond to each of the six novel sounds used in the testing phase. Each file contains a random sample of 5 different examples (copy trials), where the first example is the copy chosen for the analysis in the manuscript and the ESM material).

Supplementary Material 4: Audio Files Legends

Audio File S1 Familiar sounds (One example for each sound): 'Song', 'Blow' and 'Birdy' examples.

Audio File S2 Novel Killer Whale sounds (Five randomly chosen examples for each sound): S2.1 'Breathy Raspberry', S2.2 'Creaking Door', S2.3'Strong Raspberry', S2.4 'Wolf' and S2.5'Elephant' examples.

Audio File S3 Novel Human sounds (Five randomly chosen examples for each sound): S3.1'Hello', S3.2 'Amy', S3.3'Ah Ah', S3.4 'One Two', S3.5 'One Two Three' and S3.6 'Bye Bye' examples.

Supplementary Material 5: MATLAB scripts and Codes.

```
% model
[x1, fs] = audioread('Song Model.wav');
x1 = x1(:);
% copy
[x2, fs] = audioread('Blow Copy.wav');
x2 = x2(:);

L = fix(0.020*fs); % windows length
H = fix(L*(1- 50/100)); % hop

[v1, t1] = ComputeShortTermFeatures(x1, fs, hamming(L,'periodic'), L, H);
[v2, t2] = ComputeShortTermFeatures(x2, fs, hamming(L,'periodic'), L, H);

D = ToolComputeDistanceMatrix(v1, v2);

[p, C, d] = ToolSimpleDtw(D, 1);

imagesc(C)
hold on
for k = 1:size(p,1)-1
   line([p(k,2) p(k+1,2)], [p(k,1) p(k+1,1)], 'Color', 'k', 'Linewidth', 3)
   plot(p(k,2),p(k,1),'k.','MarkerSize',10);

% ========================================================================
%> @brief computes a feature from the audio data
%>
%> supported features are:
%>  'SpectralSTFT',
%>  'SpectralCentroid',
%>  'SpectralCrestFactor',
%>  'SpectralDecrease',
%>  'SpectralFlatness',
%>  'SpectralFlux',
%>  'SpectralKurtosis',
%>  'SpectralMfccs',
%>  'SpectralPitchChroma',
%>  'SpectralRolloff',
%>  'SpectralSkewness',
%>  'SpectralSlope',
%>  'SpectralSpread',
%>  'SpectralTonalPowerRatio',
%>  'TimeAcfCoeff',
%>  'TimeMaxAcf',
%>  'TimePeakEnvelope',
%>  'TimePredictivityRatio',
%>  'TimeRms',
%>  'TimeStd',
%>  'TimeZeroCrossingRate',
%>
%> @param cFeatureName: feature to compute, e.g. 'SpectralSkewness'
%> @param afAudioData: time domain sample data, dimension channels X samples
%> @param f_s: sample rate of audio data
%> @param afWindow: FFT window of length iBlockLength (default: hann), can be [] empty
%> @param iBlockLength: internal block length (default: 4096 samples)
%> @param iHopLength: internal hop length (default: 2048 samples)
%>
%> @retval v feature value
%> @retval t time stamp for the feature value
% ========================================================================
```

```
function [v, f, t] = ComputeShortTermFeatures (afAudioData, f_s, afWindow, iBlockLength, iHopLength)

    % set default parameters if necessary
    if (nargin < 5)
        iHopLength      = 2048;
    end
    if (nargin < 4)
        iBlockLength    = 4096;
    end

    % pre-processing: down-mixing
    if (size(afAudioData,2) > 1)
        afAudioData = mean(afAudioData,2);
    end
    % pre-processing: normalization (not necessary for many features)
    if (length(afAudioData)> 1)
        afAudioData = afAudioData/max(abs(afAudioData));
    end

    if (nargin < 3 || isempty(afWindow))
        afWindow    = hann(iBlockLength,'periodic');
    end

    % compute FFT window function
    if (length(afWindow) ~= iBlockLength)
        error('window length mismatch');
    end

    iOverlap = iBlockLength-iHopLength;
    % in the real world, we would do this block by block...
    [X, f, t] = spectrogram(afAudioData,...
                            afWindow,...
                            iOverlap,...
                            iBlockLength,...
                            f_s);
    % magnitude spectrum
    X = abs(X)*2/iBlockLength;

    PlotSignal(afAudioData, X, f_s, 10000);
    figure

    % compute features
%     [v(1,:), t] = FeatureTimeZeroCrossingRate(afAudioData, iBlockLength, iOverlap, f_s);
%     % energia
%     [v(2,:), t] = FeatureTimeEnergy(afAudioData, iBlockLength, iOverlap, f_s);
%     [v(3,:), t] = FeatureTimeEnergyEntropy(afAudioData, iBlockLength, iOverlap, f_s, 10);
%     [v(4,:)] = FeatureSpectralCentroid(X, f_s);
%     [v(5,:)] = FeatureSpectralEntropy(X, f_s, 10);
%     [v(6,:)] = FeatureSpectralFlux(X, f_s);
%     [v(7,:)] = FeatureSpectralRolloff(X, f_s, 0.9);
%     [MFCC] = FeatureSpectralMfccs(X, f_s);
%     v(8:20,:) = MFCC;
%     [v(21,:)] = PitchSpectralAcf(X,f_s);
%     [v(22,:)] = FeatureSpectralTonalPowerRatio(X, f_s);
%     [v(23,:)] = NoveltyLaroche (X, f_s);
%     [v(24,:), t] = PitchTimeAuditory(afAudioData, iBlockLength, iOverlap, f_s);
%     [HR, f0] = FeatureHarmonic(afAudioData, iBlockLength, iOverlap, f_s);
%     v(25,:) = HR;
%     v(26,:) = f0;
%     % chroma
%     Chroma = FeatureChroma(afAudioData, iBlockLength, iOverlap, f_s);
%     [v(27:38,:)] = Chroma;
```

```
%     [v(39,:), t] = FeatureTimePredictivityRatio(afAudioData, iBlockLength, iOverlap, f_s);
%     [v(40,:)] = FeatureSpectralKurtosis (X, f_s);
%     [v(41,:)] = FeatureSpectralCrestFactor (X, f_s);
%     [v(42,:)] = FeatureSpectralFlatness (X, f_s);
%     [v(43:54,:)] = FeatureSpectralPitchChroma(X, f_s);
%     [v(55,:), t] = PitchTimeAmdf(afAudioData, iBlockLength, iOverlap, f_s);

    % compute features
    % energia
    [v(1,:), t] = FeatureTimeEnergy(afAudioData, iBlockLength, iOverlap, f_s);
    % MFCC
    [MFCC] = FeatureSpectralMfccs(X, f_s);
    v(2:14,:) = MFCC;
    % Pitch Spectral ACF
    [v(15,:)] = PitchSpectralAcf(X,f_s);
    % chroma
    Chroma = FeatureChroma(afAudioData, iBlockLength, iOverlap, f_s);
    [v(16:27,:)] = Chroma;
    [v(28,:)] = FeatureSpectralKurtosis (X, f_s);
    [v(29:40,:)] = FeatureSpectralPitchChroma(X, f_s);
    [v(41,:), t] = PitchTimeAmdf(afAudioData, iBlockLength, iOverlap, f_s);
end
```

The majority of these matlab sources require the Matlab Signal Processing Toolbox installed. Several scripts (such as MFCCs and Gammatone filters) are based on implementations in Slaney''s Auditory Toolbox.