# **1** Supplementary Information

- 2 Dataset Reduction
- 3 The dataset used in this study consisted of 153 recordings containing 273 call combinations.
- 4 As two individuals has much higher counts compared to other individuals in the dataset, a
- 5 smaller sample of combinations was randomly selected for these individuals. Consequently,
- 6 the final sample size for analysis consisted of 1333 segments comprising 561 calls and 222
- 7 call combinations from 23 individuals in 16 groups (Table S1). Counts for calls and segments
- 8 is calculated from the vocal units which comprise combinations used in analysis.
- 9

10 Table S1. Counts of call combinations, calls, and segments for each individual in the original and reduced

11 datasets. Asterix denotes a change in number in the reduced dataset compared to the original dataset.

Group	Individual	Sex	Combinations		Calls		Segments	
			Original	Reduced	Original	Reduced	Original	Reduced
BWYa	MGGY	F	12	12	32	32	79	79
BWYb	MRGO	F	14	14	44	44	125	125
CPC	MOVY	F	5	5	10	10	20	20
ESC	VVOM	F	14	14	30	30	78	78
FMRa	BMYO	F	14	14	37	37	84	84
FMRa	RYMR	М	3	3	7	7	11	11
FMRb	MORP	М	4	4	11	11	16	16
GON	XF	F	6	6	15	15	36	36
GON	XM	М	3	3	6	6	10	10
JOG	GBYM	F	16	16	33	33	79	79
JOG	MOGO	М	12	12	40	40	104	104
KMO	MOOO	F	6	6	13	13	29	29
KMP	MMXBYX	F	1	1	2	2	3	3
KMP	PWMRMX	М	3	3	7	7	19	19
LHP	XF	F	5	5	13	13	31	31
MBG	OMGO	F	33	21*	80	52*	252	157*
MBG	RMYG	М	9	9	19	19	44	44
NH	RMXGMX	F	16	16	40	40	75	75
RVD	VVVM	F	64	25*	175	67*	349	133*
RVD	MXWW	М	9	9	23	23	43	43
SCL	XF	F	4	4	16	16	39	39
SCL	XM	М	2	2	8	8	16	16
SS	MXXGRY	F	18	18	36	36	102	102
Total Counts:			273	222	697	561	1644	1333

- 12
- 13
- 10
- 14



Fig. S1. Example of annotations created in Praat to define start and end boundaries of magpie vocal units on the three combinatorial levels: 1) combination, 2) call and 3) segment. In Praat, combinatorial levels were displayed

as distinct tiers indicated by the three rows at the bottom of the figure, where level is displayed by tier number

on the left and tier label on the right-hand sides. Blues lines indicate boundary start and end points for vocal

units on each tier. The highlighted section displays overlap in the transition between segment SH to segment DS in the call DS-SH-DS, resulting in the end of the SH segment and start of the DS segment being altered so that,

in analysis, representation of each vocal unit is unobstructed by the neighbouring segment.



Fig. S2: UMAP projection showing spectrograms of segments in latent space (N = 1333). Silhouette score for the segment labels = 0.213, and this is well above chance (H(2) = 748.786; p < 0.001), however the two smaller

28 classes appear to fit into other larger clusters (US = SH; HL = LH). Segment class colour code: DS (dark blue),

HL (light green), LH (teal), NL (red), SH (light purple), and US (dark purple).

- 30
- 31
- 32
- 33

## 34 UMAP Balanced Data Set

35 We conducted UMAP on a balanced dataset, to ensure there was no bias due to the

- 36 unbalanced count of spectrograms per segment class. The number of spectrograms were
- 37 reduced in three segment classes (DS, LH and SH) to reflect the number of spectrograms
- 38 within the smallest class (NL; N = 173 spectrograms). To do so, a smaller sample of
- 39 spectrograms was randomly selected for those individuals with the highest counts per
- 40 segment class (Table S2).
- 41
- 42 Table S2. Comparison of spectrogram count (per individual, per class) in the unbalanced vs balanced datasets.
- 43 Separate UMAP projections were run for the unbalanced and balanced datasets. The classes of segments shown
- 44 (DS, LH, SH) are those in which number of spectrograms were reduced to reflect the number in the smallest
- 45 segment class (NL; N = 173 spectrograms). Asterix denote cases in which spectrogram count differs in the
- 46 balanced dataset compared to the original count in the unbalanced dataset.

Group	Individual	Sex	Spectrogram Count						
			DS		LH		SH		
			Unbalanced	Balanced	Unbalanced	Balanced	Unbalanced	Balanced	
BWYa	MGGY	F	30	8*	11	11	31	10*	
BWYb	MRGO	F	41	9*	12	12	62	10*	
CPC	MOVY	F	10	8*	5	5	0	0	
ESC	VVOM	F	16	8*	15	13*	45	10*	
FMRa	BMYO	F	43	8*	13	13	18	10*	
FMRa	RYMR	М	7	7	3	3	0	0	
FMRb	MORP	М	9	8*	4	4	0	0	
GON	XF	F	9	8*	6	6	19	10*	
GON	XM	М	5	5	1	1	1	1	
JOG	GBYM	F	18	9*	16	13*	30	10*	
JOG	MOGO	М	36	8*	9	9	41	11*	
KMO	MOOO	F	13	8*	6	6	9	9	
KMP	MMXBYX	F	1	1	0	0	0	0	
KMP	PWMRMX	М	6	6	3	3	10	10	
LHP	XF	F	10	9*	5	5	14	10*	
MBG	OMGO	F	38	8*	21	13*	87	10*	
MBG	RMYG	М	16	8*	8	8	13	11*	
NH	RMXGMX	F	30	8*	11	11	15	10*	
RVD	VVVM	F	51	8*	25	13*	36	10*	
RVD	MXWW	М	22	8*	4	4	8	8	
SCL	XF	F	15	9*	5	5	9	9	
SCL	XM	М	6	6	2	2	4	4	
SS	MXXGRY	F	36	8*	18	13*	37	10*	
Total Co	ount:		468	173	203	173	490	173	

48



51 Fig. S3: Spectrograms of segments projected into UMAP latent space, where number of spectrograms per

52 segment label class is balanced (N = 173 spectrograms per class). Silhouette score for the hand labels = 0.505,

and this was well above chance (H(2) = 772.588; p < 0.001), however was only minimally better than the

54 projection using an unbalanced dataset (S = 0.481). Segment class colour code: DS (dark blue), LH (teal), NL

- 55 (red) and SH (light purple).
- 56
- 57

### 58 Individual Combinatorial Repertoire

59 We conducted separate projections on balanced segment datasets from the combinatorial

60 repertoires of two individuals (OMGO & VVVM; both female), to ensure the distinction

61 between segment class was reflected at the individual level. These two individuals were

62 chosen due to their large segment counts, compared to any other individual (Table S1). For

63 the projection on segments from VVVM, the number of spectrograms for three classes were

reduced to that count in the smallest class (NL; N = 61). For OMGO, there were too few

65 counts of NL segment spectrograms (N = 11), and as such this class was removed from the

66 dataset prior to analysis. Additionally, for OMGO the SH segment class were reduced to the

67 count of the DS segment class (N = 62), rather than to the count of the second smallest

68 segment class (LH; N = 33) to boost sample size for analysis (too low a sample size can

69 hinder the effectiveness of UMAP). For both individuals, each class reduction was conducted

- 70 via random selection of segments.
- 71

72 For the projection on segments produced by OMGO, data separated into distinct clusters (Fig. 73 S4A); this correlated highly with class labels (S = 0.652) and was significantly different from 74 a chance distribution (H(2) = 197.259; p < 0.001). There was further clustering within each 75 segment class (Fig. S4A), however, it is difficult to develop further inferences for within-76 segment class distinctions here due to the low sample size. Furthermore, for OMGO, there 77 was directionality in the transition between segment classes (Fig. S4B) that was very similar 78 to that observed when pooling all individuals (Results & Fig. 2B). Specifically, SH regularly 79 transitioned to DS, LH and SH. DS most frequently transitioned to SH, and less frequently to 80 LH. For this individual, the probability for a DS-to-DS transition (0.26 probability) was much 81 higher than that probability observed when pooling all individuals (0.07 probability; see 82 Results). Additionally, LH was not observed to transition to any segment, which differs to the 83 results for all individuals whereby LH was found to transition to D. This distinction, 84 however, is expected due to this individual being located in the Crawley study site, wherein 85 calls containing the LH to DS transition are rare. Overall, these results largely support the 86 distinctions between segment classes and directionality in probability for segment transition, 87 that was shown in the projection pooling all individuals together (Results & Fig. 2). 88



89

Fig. S4. A: Spectrograms of segments produced by the individual OMGO projected into UMAP latent space (N
= 157). The legend in S4B can be applied to similarly coloured data in S4A. B: Data are coloured by segment
class label (DS: dark blue; LH: teal; SH: light purple). Transitions between segment classes (symbolized by
coloured circles). Arrowed lines represent transitions between segments with opacity increasing in relation to

94 transition probability (value indicated by text next to respective lines).

95

Likewise, for the projection on segments produced by VVVM, data separated into distinct 96 97 clusters (Fig. S5A); this correlated highly with class labels (S = 0.548) and was significantly 98 different from a chance distribution (H(2) = 198.864; p < 0.001). Moreover, directionality in 99 segment transition (Fig. S5B) was very similar to that observed when pooling all individuals 100 (Results & Fig. 2B). Specifically, SH was found to transition to DS, LH and SH, which is 101 very similar to the transitions observed for all individuals whereby transition from SH to NL was very rare (0.01 probability; Fig. 2B). DS most frequently transitioned to SH (0.8 102 103 probability), less frequently to LH (0.2 probability), and never to DS or NL. These echo the 104 findings for DS segment transition that were observed when pooling all individuals, whereby 105 there is very low probability for a DS-to-DS transition (0.07 probability) and there is a 106 probability of zero for a transition to NL (see Results & Fig. 2B). Additionally, LH only 107 transitioned to DS (probability of 1) which mirrors the respective transition probability that 108 was observed for the projection on all individuals. On the whole, these findings support those 109 results for segment class distinction and directionality in sequential transition found in the 110 analyses on all individuals (Results & Fig. 2).



113 Fig. S5. A: Spectrograms of segments produced by the individual VVVM projected into UMAP latent space (N

114 = 244). The legend in S4B can be applied to similarly coloured data in S4A. B: Data are coloured by segment

115 class label (DS: dark blue; LH: teal; NL: red; SH: light purple). Transitions between segment classes

116 (symbolized by coloured circles). Arrowed lines represent transitions between segments with opacity increasing

- 117 in relation to transition probability (value indicated by text next to respective lines).
- 118

## 119 Within-Segment Class Analysis

120 Table S3. Spectrogram count for each within-segment class UMAP projection. For the SH segment class

- 121 analysis, two individuals had much higher counts of spectrograms compared to any other individual. As such,
- 122 for those individuals, a smaller portion of spectrograms was randomly selected to be included in the analysis.
- 123 Asterix denotes a change in the number of spectrograms in the reduced dataset, compared to the original dataset.

Group	Individual	Sex	DS	LH	NL	SH	
						Original	Reduced
BWYa	MGGY	F	30	11	7	31	31
BWYb	MRGO	F	41	12	10	62	43*
CPC	MOVY	F	10	5	5	0	0
ESC	VVOM	F	16	15	1	45	45
FMRa	BMYO	F	43	13	10	18	18
FMRa	RYMR	М	7	3	1	0	0
FMRb	MORP	М	9	4	3	0	0
GON	XF	F	9	6	2	19	19
GON	XM	М	5	1	3	1	1
JOG	GBYM	F	18	16	15	30	30
JOG	MOGO	М	36	9	18	41	41
КМО	MOOO	F	13	6	1	9	9
KMP	MMXBYX	F	1	0	2	0	0
КМР	PWMRMX	М	6	3	0	10	10
LHP	XF	F	10	5	2	14	14
MBG	OMGO	F	38	21	11	87	44*
MBG	RMYG	М	16	8	7	13	13
NH	RMXGMX	F	30	11	19	15	15
RVD	VVVM	F	51	4	9	36	36
RVD	MXWW	М	22	25	22	8	8
SCL	XF	F	15	5	10	9	9
SCL	XM	М	6	2	4	4	4
SS	MXXGRY	F	37	18	11	37	37
Total Count:			468	203	173	490	427*

- ----





135 136 Fig. S6: Separate projections for each segment class; data is coloured by study site. A: LH segments projected 137 into UMAP latent space (N = 203). Study site partly described distribution of data across the projection (S = 138 0.097), and was significantly different from a random distribution (H(2) = 19.724; p < 0.001). B: UMAP on NL 139 segments (N = 173). NL segments formed two distinct clusters and study site marginally described distinction 140 here (S = 0.028), but was not significantly different from chance (H(2) = 1.836; p = 0.175). C: UMAP on SH 141 segments (N = 427). There was some clustering of SH segments, which may be partly correlated with study site 142 (S = 0.093), and was significantly different from a random distribution (H(2) = 104.702; p < 0.001). Clusters for 143 each site, however, were not well defined or separated. D: UMAP on DS segments (N = 468), showing two 144 distinct clusters. Study site marginally explained distribution of spectrograms (S = 0.036), and was significantly 145 different from a random distribution (H(2) = 33.719; p < 0.001). However, as spectrograms from each site were

146 found across both clusters, this predictor did not appear to reflect the true distinction between clusters.

For segments produced by VVVM, there was further clustering within each segment class 148 149 (Fig. S5A & S7) and, for two segment classes, these mirrored findings for within-segment 150 class variation when pooling all individuals (Results & Fig. 3). Specifically, spectrograms 151 within the LH and NL segment classes clustered according to their position, relative to 152 arrangement with neighbouring segments. For one cluster, all bar one of the spectrograms are 153 lone NL segments (the segment itself making up a call-level unit), whereas another smaller 154 cluster contained only NL segments produced in combination with other segments (Fig. S7). 155 Furthermore, LH segments formed four clusters with two groupings: i) those that precede a 156 DS segment, and ii) those that are positioned at the end of a call. It should be noted that there were two clusters for each grouping of LH segments, with one small cluster from each 157 158 grouping being located closer to a cluster of SH segments, rather than their respective LH 159 grouping. As duration is the main difference between LH and SH segments, these cases of distinction within the LH groupings may suggest an intermediate classification, wherein 160 161 segments possess a duration between that of SH and other LH segments. However, due to the low sample size, this classification cannot be validated. Overall, results closely reflect that 162 163 shown in the projection using all individuals.



166

167 Fig. S7: Spectrograms of segments produced by the individual VVVM projected into UMAP latent space (N =

168 244). For DS and SH segments classes, data are coloured as in Fig. 2 (see Results; DS: dark blue, SH: light

169 purple). For LH and NL segment classes, data are coloured by position in relation to neighbouring segments

170 comprising the same call (as in Fig 3, see Results). Namely, LH segments positioned at call-end (light green),

171 versus LH segments preceding a DS segment (teal), and NL segments produced in isolation (separated by

172 silence; red) versus those in combination with other segments within a call (light orange).

- 173
- 174

#### 175 Analysis of within-call group variation

176 We conducted further UMAP analyses to investigate whether there was further separation of 177 call types within each of the three broad classes. As such, we ran three projections on the 178 following: i) all call spectrograms containing an LH segment, ii) calls comprising any lone 179 DS, NL or SH segment, or those segments in any combination, and iii) calls comprising DS 180 or SH lone segments, or any combination of DS, NL and/or SH segments. Because many 181 calls contained segments that were repeated, we used a simplified call labelling system such 182 that labelling excluded display of repeated consecutive segments. For example, a call 183 comprised of three SH segments followed by a LH segment would be given the label SH-LH. 184 Per projection, we created labels for groupings of similar call types, based on associations 185 found in the distribution of spectrograms in the model (see below for further detail). 186 Inferences from these results were limited due to the overlap or closeness of clusters in the 187 projections, however, are presented here as they illustrate there is some extent of separation 188 in call spectrograms, which may be bolstered in future work using a larger dataset. 189 190 Looking at the projection of spectrograms of any call containing an LH segment (N=203), the 191 silhouette score did not suggest correlation with simplified call labelling (S = -0.039; Fig. S8A), and distribution was only just significantly different from random (H(2) = 4.728; p = 192 193 0.030). The calls, however, appeared to lie along a continuum in which there appeared to be 194 two pairs and one trio of similar call labels that occupy slightly overlapping regions (Fig. 195 S7B). Firstly, calls containing an LH segment followed by a DS segment (LH-DS and SH-196 LH-DS calls) appeared to cluster slightly apart from other LH segment containing calls. 197 Secondly, LH lone segment calls clustered near SH-LH calls. Finally, there was a cluster of 198 calls characterized by having a DS segment that was arranged at any position within the call 199 prior to an LH segment (SH-DS-LH, DS-SH-DS-SH-LH, and SH-DS-SH-LH calls). 200 Although there did appear to be some gradation between these three groups of similar calls, 201 there was relatively little overlap (Fig. S8B). The minimal overlap between the three groups 202 suggests these may reflect distinct call types, which is supported by the silhouette score (S =203 (0.327) and this was significantly different from random (H(2) = 81.846; p < 0.001). 204



Fig. S8: Spectrograms of any call that contains an LH segment projection into UMAP latent space (N = 203). A: Simplified call type was found to inadequately describe the distribution of data across the projection (S = 0.039)

and was only just significantly different from random (H(2) = 4.727; p = 0.03). Data is coloured by simplified

- 209 call label. B: The three groupings of similar call labels described the distribution of data across clusters quite
- 210 well (S = 0.327) and was significantly different from chance (H(2) = 81.846; p < 0.001). Data is coloured by the
- 211 three groupings of similar call labels. C: Spectrograms from the two study sites occupy largely separate regions
- 212 within the model, suggesting there is an acoustic difference in call spectrograms between the two sites.
- 213 Silhouette score shows study site explained part of the distinction between spectrograms (S = 0.147) and this
- 214 was significantly different from a random distribution (H(2) = 38.920; p < 0.001). Data is coloured by study site.
- 215
- 216 Furthermore, there appeared to be a correlation between spectrogram distribution and study
- site, with data from each site occupying largely separate regions in the model (Fig. S8C).
- 218 Despite minimal overlap of data between study sites, this was not prominently supported by

219 the silhouette score (S = 0.147), though was significantly different from a random distribution (H(2) = 38.920; p < 0.001). However, spectrograms from the Guildford study site appeared to 220 221 cluster into two groups at opposite poles of the data continuum in the model, which was 222 likely the reason for the lower silhouette score. Interestingly, in comparing the distribution of 223 data when coloured for study site versus when the data is coloured for similar call labels, 224 there appeared to be a relationship. Specifically, calls comprising an LH segment either with 225 or without a preceding SH segment appeared to predominantly occur in the Crawley site, 226 whereas the Guildford site appeared to hold most of the call spectrograms from the two other 227 groupings of similar calls (Fig. S8B-C). This suggests that there may be some location 228 differences in the types of calls that are produced between the two study sites. No other

229 predictor was able to describe distribution of data across the projection (S < 0).

230

231 For the projection of calls that comprise any lone DS, NL or SH, or contain those segments in 232 any combination, the silhouette score did not suggest correlation with simplified call labelling (N = 358; S = -0.280; Fig. S9A), and distribution was not significantly different from random 233 (H(2) = 0.647; p = 0.421). NL segment calls, however, appeared to cluster separate from all 234 235 other calls, suggesting they vary acoustically compared to the other calls and may comprise a 236 distinct call type. As such, we ran another projection excluding any NL segment only calls (N 237 = 332; Fig. S9B). Again, we found no correlation between the distribution of spectrograms 238 across the projection and the simplified call label (S = -0.328) and this was not significantly 239 different from a random distribution (H(2) = 2.440; p = 0.118). There did, however, appear to 240 be some level of aggregation, albeit disperse, for groupings of similar call labels. 241 Specifically, there appeared to be three diffuse clusters: i) NL-DS calls, or any combination 242 of NL with DS and SH segments, ii) lone DS segment calls, and iii) lone SH segment calls or

any combination of DS and SH segments (Fig. S9C). Silhouette score indicated this way of

grouping the calls marginally described distinctions between spectrograms (S = 0.099) and

this was significantly different from random (H(2) = 40.405; p < 0.001). Consequently, there

is suggestion that spectrograms here may have clustered into distinct call types (according to groupings of similar call labels), however, it was difficult to draw conclusions here due to the

248 diffuse and overlapping nature of the clusters in the model. No other predicter could better

explain distribution of spectrograms across the projection (S < 0.067).





Fig. S9. Projections for spectrograms of calls containing the segments DS, NL or SH (lone or in any

253 combinatorial arrangement). NL lone segment calls are excluded from the analysis in B-C. A: Data coloured by

simplified call label which was found to inadequately describe distribution of data across the projection (S = -

255 0.280; N = 358) and was not significantly different from a random distribution (H(2) = 0.647; p = 0.421). B:

256 Data coloured by simplified call label (N = 332), which did not correlate with distribution of data (S = -0.328)

and was not significantly different from random (H(2) = 2.440; p = 0.118). C: Data coloured by groupings of similar call labels (N=332): i) NL-DS calls or any combination of NL with DS and SH segments, ii) lone DS

- segment calls, and iii) lone SH segment calls or any combination of DS and SH segments. Grouping calls in this
- 260 manner partly described the distribution of spectrograms across the projection (S = 0.099) and this was
- 261 significantly different from random (H(2) = 40.405; p < 0.001).