

1 **Phylogenetic inference for function-valued traits: speech sound**
2 **evolution**

3
4 **The Functional Phylogenies Group**

5
6 **Corresponding Author: Nick Jones nicholas.jones1@imperial.ac.uk**

7 **John A.D. Aston - Warwick University, Statistics, UK**

8 **Dorothy Buck - Imperial College London, Mathematics, UK**

9 **John Coleman - Oxford University, Phonetics Laboratory, UK**

10 **Colin J. Cotter - Imperial College London, Aeronautics, UK**

11 **Nick S. Jones – Imperial College London, Mathematics, UK**

12 **Vincent Macaulay - Glasgow University, Mathematics and Statistics, UK**

13 **Norman MacLeod – Natural History Museum, Palaeontology, UK**

14 **John M. Moriarty - Manchester University, Mathematics, UK**

15 **Andrew Nevins - University College London, Linguistics]**

16
17 **Abstract:** Phylogenetic models have recently been proposed for data that are best
18 represented as a mathematical function (function-valued). Such methods can used
19 to model the change over time in function-based descriptions of various data of
20 interest to evolutionary biologists, including the sound of speech. This approach to
21 phylogenetic inference and analysis is challenging, both in terms of modelling the
22 phylogenetics of functions, and in engaging with previously existing evidence for
23 character-state change. Nevertheless, it is both a real and exciting prospect. Our
24 approach could provide those interested in investigating a greater range of

25 evolutionary processes with the ability to utilize statistical hypothesis testing
26 procedures and to create estimates of the states of function-valued characteristics
27 (e.g., speech sounds) at earlier historical times.

28

29 **Beyond symbols**

30

31 The word 'functional' has multiple meanings. One of these — and the one we will
32 focus on in this article — is 'represented by a continuous mathematical function'.

33 Given a set of traits that can be described as functions (these are called function-
34 valued traits [1]) and that have been mapped to positions on a network of

35 phylogenetic relationships, what can be said about values of these traits at other
36 points on the phylogeny? While the study of the evolution of both genetic

37 sequences and quantitative characters, in particular comparative method analysis,
38 has seen extensive application and technical development, the study of the

39 phylogenetics of function-valued characteristics is comparatively less advanced.

40 Herein we propose methods that could be useful in extending the ability of

41 evolutionary inferences in many areas of evolutionary system studies. Function-

42 valued traits are quite common and appear across many areas of biology, from

43 human height vs age to backfat thickness of animals vs weight. However, the idea

44 that traits can be represented as functions is new to other fields less closely

45 associated with biodiversity studies. One such field is linguistics. Accordingly,

46 speech sound evolution will serve as a running example of our proposed approach.

47 As background to this choice we note that this is a time of great promise in linguistic

48 cladistics. For example, phylogenetic methods have been used to shed light on the

49 early differentiation of Indo-European languages and the dynamics of the

50 appearance of new languages [2,3]. There have also been agent-based simulations
51 investigating how speech sounds might evolve, e.g. [4].

52

53 One method for dealing with function-valued traits in a phylogenetic context is the
54 functional phylogenetic regression method recently proposed by Jones and Moriarty
55 [5]. This data analysis technique differs from previous approaches in its use of
56 continuous characters [6,7,8] and by modelling evolved traits explicitly as functions.
57 Building upon recent work in machine learning [9], this mode of phylogenetic
58 regression takes account of constraints on the patterns of variation for function-
59 valued traits over a phylogeny that are not captured by more traditional multivariate
60 models. Since this approach avoids summarizing these data it also supports the
61 prediction of ancestral function forms directly rather than through the estimation of a
62 few of their characters. For speech-sound analysis we could model each articulatory
63 movement as a sequence of positions of the vocal apparatus, a continuous view of
64 which was achieved in a straightforward manner. In continuing our efforts to develop
65 this approach our object for enquiry is simply the speech sound itself (or functional
66 versions of it such as the spectrogram or empirical covariance surface [see
67 Glossary]) which can be treated as a mathematical function.

68

69 Our primary goal is to discuss how quantitative methods can allow us to infer
70 probability distributions over possible ancestral function-valued traits (see Fig. 3).
71 Samples from these distributions offer the prospect of estimating functional
72 descriptions of speech from languages which have no living speakers. Provided this
73 prospect can be realized rigorously, such an ability would be a substantial advance
74 from symbolic phylogenetic modelling towards (in this case) true acoustic

75 reconstruction, yielding a new tool in comparative philology. In a larger sense,
76 however, this function-based approach has relevance for many types of biological
77 data and, indeed, for phylogeny reconstruction itself.

78

79 The approaches proposed below lend themselves to testing hypotheses about the
80 sounds of past languages suggested, for example, by textual inference or from the
81 written descriptions of phoneticians in the ancient world e.g. [10,11,12,13]. Our
82 attempts to realize this aim are of course quite preliminary. Nevertheless, we are
83 intrigued by the prospect of using phylogenetic methods to enable us to hear
84 approximations to the sound of dead languages. This goal is most naturally
85 approached through the analysis of mathematical functions describing actual
86 speech. Fortunately, we are close to having all the materials and mathematical
87 methods needed to make the first attempts toward this goal's realization.

88

89 **Speech sounds as function-valued traits**

90

91 Speech sounds have a wide range of component frequencies, from a few hertz (e.g.,
92 speaking vs. silence) to c. 6 kHz, with a dynamic range of c. 60 dB. Despite this
93 variability, speech sound waves are fairly similar over time intervals of 5-10 ms or
94 more, the higher frequencies arising from multiples of the fundamental frequency
95 (see Glossary) in voiced sounds, and resonances of the vocal tract. Therefore, it is
96 possible - indeed, now routine in many areas of speech processing - to decompose
97 speech sound waves into time series of multiple (c. 12-20), slowly-varying
98 components which are approximately constant on a timescale of 5-10ms. As a result,
99 the representation of speech sound as a function, a function-valued trait, or set of

100 functions is common within phonetics and speech technology e.g. in Linear
101 Prediction Coding [14].

102

103 **Words as functions**

104

105 Speech sounds have far-reaching contextual effects on one another (= co-
106 articulation [15]). It is therefore very difficult to decompose recordings of speech into
107 separate, consecutive 'sounds' with a high degree of objectivity and repeatability.
108 Although symbolic representations such as phonetic transcriptions are based on
109 separate symbols, in the continuum of articulated speech, sounds merge into one
110 another [16,17]. As we will discuss below, historical and evolutionary sound changes
111 rarely affect individual component sounds. Rather, sounds evolve together within the
112 context of specific words. Accordingly, we may sidestep the practical difficulties of
113 segmenting words into separate sounds by modelling the evolution of (sets of) whole
114 words. The challenge of segmenting vocalizations becomes even more pronounced
115 when considering animal sounds e.g. [19,20,21]. A functional approach to the
116 analysis of these data is not only the most appropriate for the analysis of human
117 speech sounds, it is also the most generalisable.

118

119 **Can speech sound change can be regarded as incremental?**

120

121 Language change is the result of imperfect learning, reanalysis, gradual changes in
122 pronunciation, and/or recurrent misperception of particular sequences [see e.g. 22].
123 Consider the differences between Spanish, Catalan and Portuguese, three
124 languages that originate from a common ancestor (Iberian, itself descended from

125 Latin). For example, the *bl-* sound is observed in two of the three languages (Sp.
126 *blanco*, Cat. *blanc*) and *br-* in one of them (Port. *branco*), as well as *bl-* being in the
127 original ancestor of all three. This change, from *l* -> *r*, is necessarily context-
128 sensitive. It did not occur across-the-board in these languages — in fact, it never
129 occurs at the beginning of a word (cf. Sp. *libro*, Cat. *llibre*, Port. *livro*), but only after a
130 preceding consonant in the same syllable. This change pertains specifically to the
131 similarity of *bla-* to *bra-* in two crucial regards: first, it shows that the two are similar
132 enough that such a change is gradual and minimal in terms of the articulators
133 involved and second, in the perceived acoustic similarity, rather than a wholesale
134 change of, say, *bla-* to *shroo-*. The comparison between words beginning with *bl-* vs.
135 *br-* in different branches of the Iberian languages represents the result of
136 accumulated and recurrent sound changes reflecting the long-term dynamics of
137 repeated listener/speaker mislearning, misperception, mispronunciation and
138 cognitive reorganization.

139

140 Rather than occurring at the level of single consonants and vowels (e.g., *l* changes
141 to *r*), could it be the case that continuous changes to the entire word are a better way
142 of understanding speech sound change? Consider one of the most common
143 linguistic changes, namely substitution. While it is fruitful to formalize the patterns of
144 change in a language as operations on strings, symbolic rules such as $xyz \rightarrow xzz$ do
145 not always capture the fact that sound changes are non-arbitrary: sound substitutes
146 are similar sounds. For example, Latin *clavis* (with initial [k-]) became Italian *chiave*
147 [kjave]; *civitas*, (also with initial [k-]), became Italian *citta* ([tʃitta]) and French *cit *, with
148 initial [s] . Numerous examples of this chain of sound changes (in the phonetic
149 alphabet $k \rightarrow kj \rightarrow cç \rightarrow tʃ \rightarrow ts \rightarrow s$) can be found across many language families.

150 **A functional perspective on sound change**

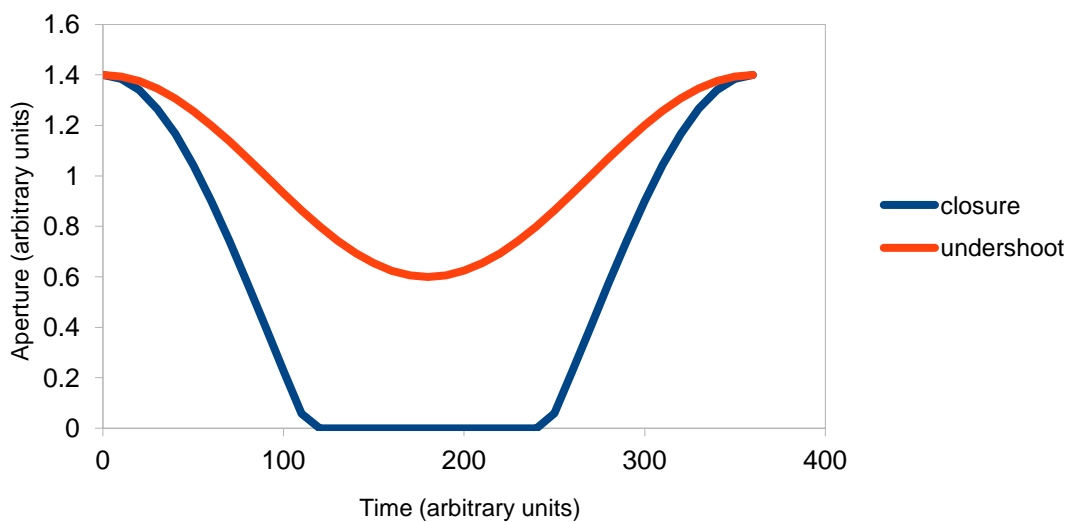
151

152 To understand such historical processes, it is necessary to make the notion of

153 'similarity' more precise. Two articulatory reasons for substitutions are undershoot

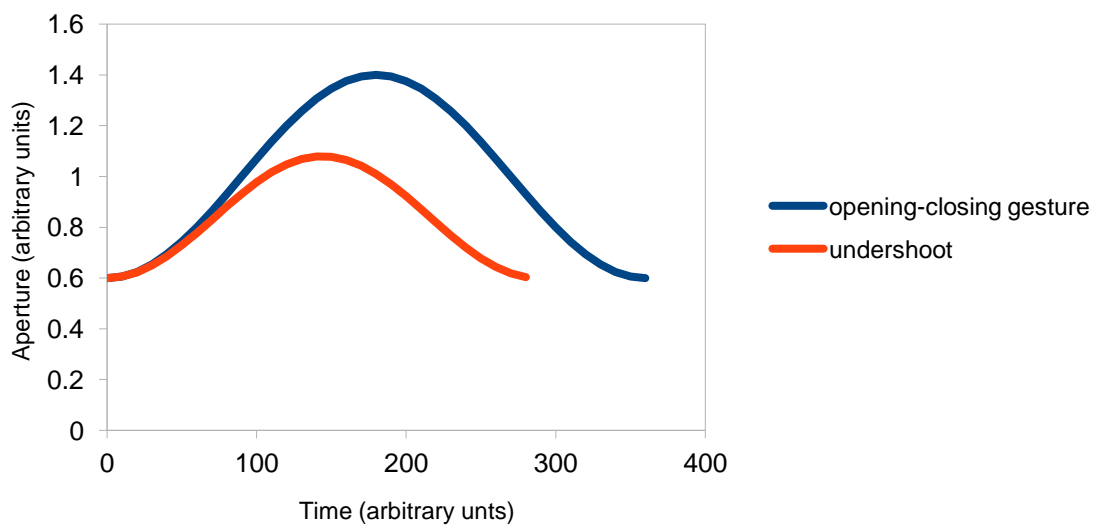
154 and overshoot (see Glossary and Fig. 1A and 1B).

Figure 1A



155

1B



156

157 Figure 1A. Here the y-axis is the size of the aperture of the oral airway, i.e. of the vocal tract
158 **Undershooting a closing movement:** e.g. [t] → [s], the 't' in *permit* becomes the 's' in
159 *permission* (noncontinuant sounds (stops) become continuants) or Latin *spatium* becomes
160 French *espace* and English *space*. Blue line: An articulatory movement that completely
161 closes the vocal tract (aperture = 0 from t= 100 to t = 250), as for [t] . Red line: An
162 incomplete closing movement, as for [s]. The vocal tract is constricted, but before complete
163 closure is reached, the tract re-opens. 1B. **Undershooting an opening movement:** e.g. a
164 low (open) vowel becomes less wide open, as in *telegr[a]ph* → *telegr[ə]phy* [a] → [ə]. Blue:
165 An opening-closing movement of the vocal tract, as for [a]. Red: An incompletely executed
166 opening movement, as for [ə]. Perhaps under a time constraint, the vocal tract begins to
167 close again at t = 150, before it has reached the target aperture.

168

169 The two other types of sound change, deletion and insertion, can also be understood
170 in terms of functions. As an operation on strings, deletions are of the form ... axb ...
171 → ... a b ... but, viewed physically and continuously, we can consider deletion to be
172 the gradual shortening of x until its duration becomes negligible. For example,
173 progressive shortening of vowels to the point at which they have no duration gives
174 reduced pronunciations (e.g., *suppose* → s'p'pose). The ubiquitous redundancy in
175 speech allows deletion to occur without significant loss of information. Viewed as an
176 operation on strings, the addition or insertion of sounds has them appear from
177 nowhere: ... a b ... → ... a y b From a functional perspective inserted sounds are
178 often comprehensible in terms of their context and can be sometimes associated
179 with a misphasing phenomenon.

180

181 We have thus seen that we can reinterpret attested sound changes – deletions,
182 insertions, and substitutions – in terms of continuous changes to continuous
183 functions, rather than discrete and arbitrary transformations of strings.

184

185 **Analysis of ‘functionalized’ speech sounds**

186

187 Having argued that speech sounds and their evolution can be treated in terms of
188 mathematical functions, we are now in a position to describe the tools that enable us
189 to use mathematical functions as characters in a phylogenetic analysis.

190

Box 1 Gaussian Process Regression

Examples of Gaussian processes are the Brownian motion and the Ornstein-Uhlenbeck processes [9] (see Fig. 2A). The latter can be viewed, informally, as a Brownian motion which tends towards a fixed mean. Both Brownian motion and Ornstein-Uhlenbeck processes have been used to describe and analyse continuous characters as part of the comparative method in phylogenetics [6-8]; they are used as models of how univariate continuous traits might evolve in time. It is worth noting that, while such continuous traits are normally used in approaches that assume a fixed phylogeny (e.g., the comparative method), they can, in principle, be used for phylogenetic inference. In this contribution we are interested primarily in how function-valued data (not only univariate data) evolve in time.

Linear regression fits a function $a.l + b$ (where a and b are fitting parameters) to N data values $\{y(l_1), y(l_2), y(l_3) \dots y(l_N)\}$ measured at some N co-ordinates of observation $\{l_1, l_2, l_3 \dots l_N\}$. Gaussian process models can also be used to fit function-

valued data. In a peculiar fashion, Gaussian process regression presumes a fit of a multivariate normal or Gaussian distribution to data: a multivariate Gaussian where each co-ordinate of observation in a set of N such measurements, $\{l_1, l_2, l_3 \dots l_N\}$, is assigned a variable in the N -dimensional distribution. The data values, y , at these points, $\{y(l_1), y(l_2), y(l_3) \dots y(l_N)\}$ are just one sample from this distribution. Just like any multivariate Gaussian these processes are specified by how the variables co-vary (the matrix of covariances): different types of covariances, or couplings, between the co-ordinates of observation yield different process models. In Figure 2A the points on the blue curve covary differently from points on the black curve. Gaussian process models are called non-parametric because, unlike parameterized models (like $a \cdot l + b$) the number of parameters increases with the number of observations N . A final relevant property of Gaussian processes is that they can be used, given knowledge of data values at some co-ordinates of observation $\{y(l_1), y(l_2), y(l_3) \dots y(l_N)\}$, to make predictions about values at other co-ordinates $\{r_1, r_2, r_3 \dots\}$ about which we do not have any data (see Fig. 2b). Box II considers a generalisation of this prediction task to unmeasured data at different evolutionary times.

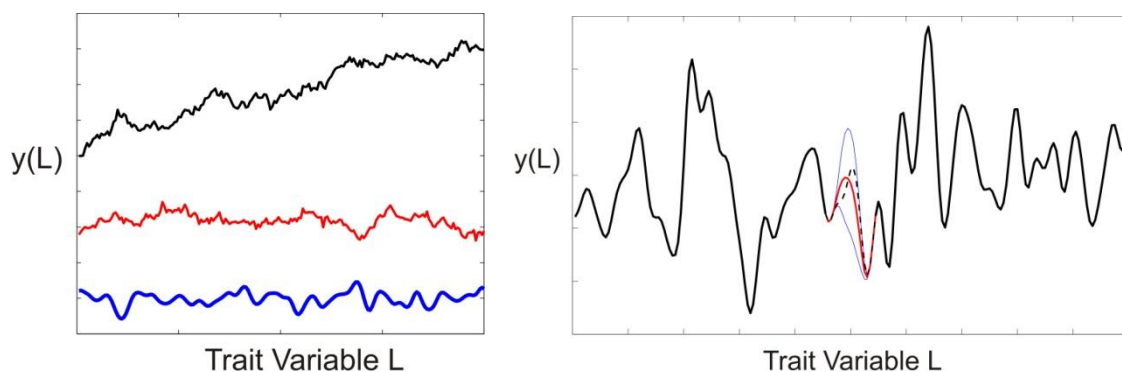


Figure 2. (A) Example Gaussian Process samples - Black line: a sample from a Brownian Motion process. Red line: a sample from an Ornstein-Uhlenbeck process. Blue line: a sample from a process with a squared-exponential covariance function.

Processes like these can be fit to data easily. (B) Making inferences about unseen data within a function-valued trait: A missing data problem. The data are represented by the black curve (we assume the data were generated by a Gaussian process). The red curve represents the locus of expected values at the missing data points and the blue curves are the one standard deviation uncertainties at each value. The black-dashed curve is a sample from the process during this period of missing data. In Box 2 we will treat unseen ancestral functions as missing data.

191

192 The treatment of data as discretely observed mathematical functions or curves has
193 received substantial attention in the recent mathematical, statistical and machine
194 learning literature. In Functional Data Analysis [23,24] the data are often assumed to
195 arise from a stochastic process which yields smooth functions (possibly subject to
196 noise). Again, note that the word 'functional' here, and in the following, refers to
197 mathematical functions, not an aspect of biological function. Considerable progress
198 has been made in extending standard statistical notions that are in common use for
199 univariate or multivariate data (e.g., regression, classification) to these functional
200 data objects [25]. Box 1 outlines one type of functional data object: a Gaussian
201 process. Further, recent work in quantitative genetics has applied functional data
202 analysis to function-valued traits [26,27]. However, little work has been done to date
203 using functional data in phylogenetic inference, particularly in a non-parametric
204 context.

205

206 Explicitly functional techniques have been applied in linguistic analysis. Some of the
207 earliest research on functional data analysis corresponded to tracking articulatory lip
208 movement trajectories during speech [28]. A recent empirical study has shown that,

209 in carefully controlled conditions, Functional Data Analysis can help classify different
210 yet closely related consonant sounds [29]. Some researchers are now taking a
211 functional approach to word contractions like the suppose→s'ppose example
212 discussed above e.g. [30]. Several studies have considered functional
213 representations of f_0 curves (f_0 , the fundamental frequency, is the acoustic basis of
214 perceived voice pitch; see Glossary). Grabe et al. [31] used a component set of
215 functions to examine f_0 intonation contours. Changes in intonation (tones) which can
216 impart meaning to words are present in many of the worlds' languages, including
217 languages such as Mandarin Chinese. Non-parametric functional data analysis has
218 been used [32] to characterise the tones of Qiang, a Sino-Tibetan language. It was
219 shown, first, that known tonal f_0 patterns can be reproduced from the empirical
220 covariance surface (see Glossary) and, second, that additional patterns, not
221 necessarily postulated from isolated utterances can be identified as relevant. This
222 was then found to apply to a much larger body of f_0 curves in Mandarin [33].

223

224 **Evolutionary Inference for function-valued data**

225

226 The phylogenetic Gaussian process models that we have outlined in boxes 1 and 2
227 describe one way in which we could perform evolutionary inference with function-
228 valued data. Figure 3 provides an example of attempting to predict an ancestral
229 value (which in the case of speech, can be tested against our other knowledge of
230 past speakers, see below). But we note that one can also use this approach to
231 support direct phylogenetic inference. Gaussian processes make a minimal set of
232 assumptions about the patterns which they model. These assumptions are that the
233 distribution of each value of each trait is Gaussian, and all correlations between time-

234 points in the signal are specified only by how pairs of points covary (the data is
235 assumed to have only pairwise correlations). Though these might seem to be strong
236 constraints, the Gaussian process framework is remarkably flexible and robust when
237 applied to a wide variety of inference tasks [9].
238

Box 2 Gaussian Processes for Evolutionary Inference

In Box 1 we noted that a Gaussian process's nature (just like any multi-variate Gaussian) is specified by how the co-ordinates of observation, $\{l_1, l_2, l_3, \dots, l_N\}$, covary. We have given examples where the set of co-ordinates of observations corresponds to observations or measurements at successive points in time. In fact, these points of observation need not only be in one dimension: they could be observation co-ordinates of a 2D function on a plane (e.g., the x - y co-ordinates on a surveyor's topographical map).

Let us suppose we call one dimension evolutionary time and the other dimension trait-space. A 2D function can then be viewed as a 1D function with spatial co-ordinates (a function-valued trait) evolving in time. If we think of the curve changing its form down an evolutionary tree (see Figure 3), any path from tree root to branch tip can be viewed as a 2D function. Jones and Moriarty [5] take this intuition and work out the correct form for the covariances between points of observation anywhere on any given tree T provided we know — or can assume — how co-ordinates of observation on the path from root to tip covary. The degree of covariance depends on the phylogeny. Different phylogenies, T_1 and T_2 , will yield different co-variances because the relative positions of co-ordinates of observation, $\{l_1, l_2, l_3, \dots, l_N\}$, will

change. Given knowledge of the covariances one can formulate likelihoods for the different tree topologies, calculate appropriate Bayes factors, and so determine which phylogeny, T_1 or T_2 , is most consistent with the observed data.

Once we understand how traits at different points on the tree covary, we can also develop a distribution for trait values at internal points and/or nodes on the tree Fig. 3. For example, using such a distribution it would be possible to attempt to forecast trait values at any points in the tree with co-ordinates $\{r_1, r_2, r_3, \dots\}$, given only observations at the tree's tips $\{y(l_1), y(l_2), y(l_3) \dots y(l_N)\}$ in exactly the same way as we approached the missing data problem in Fig 2B. It is also possible to obtain a sample from a distribution for trait values at internal points to yield example ancestral functions.

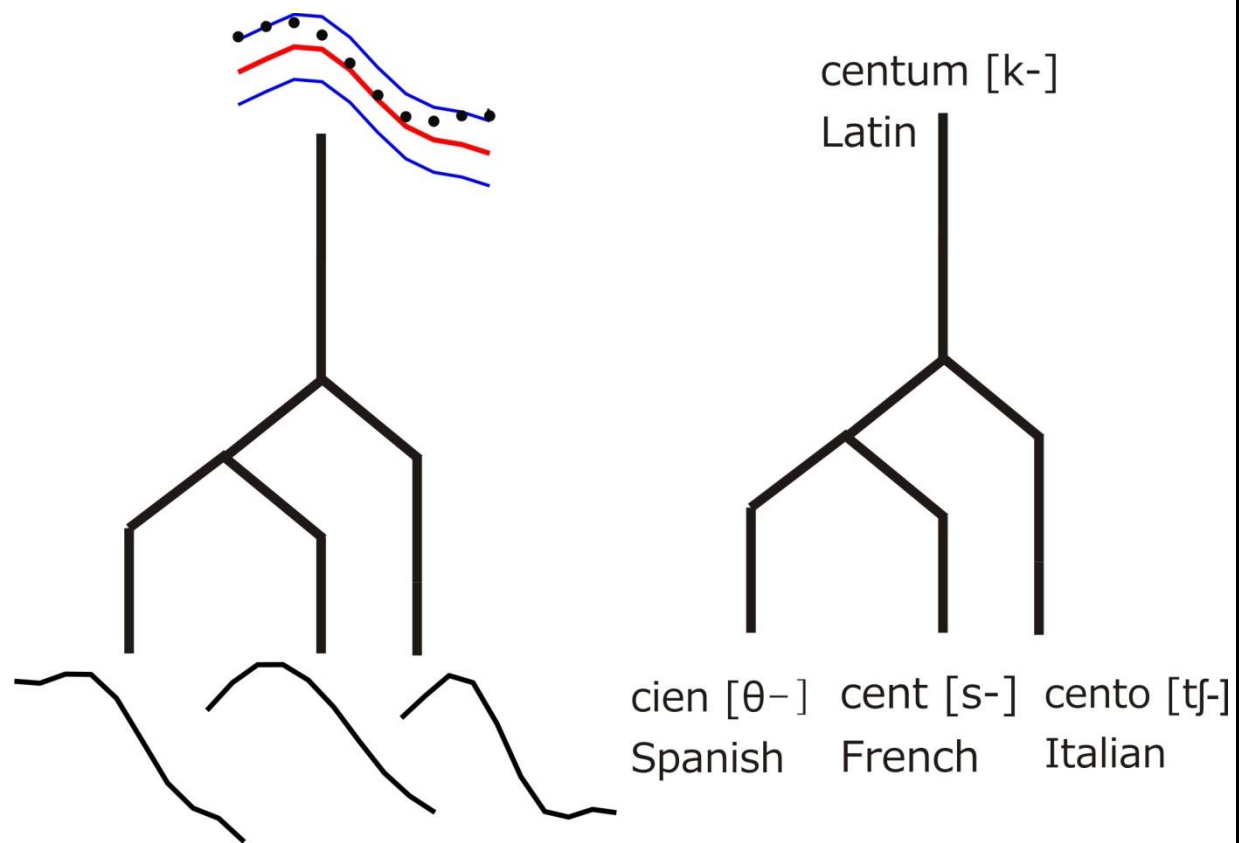


Fig. 3. *Making inferences about unseen data pertaining to an ancestral function-valued trait: Phylogenetic Gaussian process regression.* The black lines are the function-valued data we observe at the leaves of a known evolutionary tree and we assume a notional curve in the deep past which is flat. Two curves have a common ancestor at two units of evolutionary time and all three have a common ancestor at three units of evolutionary time. We construct a probability distribution over function-valued traits at 5 units of evolutionary time (we do this by modelling the data as having a covariance in the space of the function-valued trait like that of the smooth blue curve in Fig. 2a and having an Ornstein-Uhlenbeck covariance in evolutionary time, like the red curve in 2a and we also allow a small amount of noise). The red line is the expected value of the ancestral function at each point in the trait space and the blue lines are one standard deviation uncertainties in that value. The black dots are a sample of the Gaussian process at the earlier time, conditional on the data at the tips. Note that we could alternatively have attempted to infer the phylogeny most consistent with the function-valued tip data. We have placed a notional example next to this for a speech sound evolution.

239

240 The use of the full Gaussian process framework described in the boxes is just one
241 route to evolutionary inference with function-valued data. Another option would be to
242 pool each piece of functional data available and use these together to construct a
243 small set of component functions that are appropriate for representing any member
244 of the set (e.g., obtaining a set of basis functions by using a functional version of
245 principal components analysis [23,24]). This would yield a natural set of functions
246 with which to decompose any one of the function-valued traits; i.e., any function can
247 then be viewed as a weighted combination of these component functions. As a trait
248 varies through evolutionary time it could be that this set of component functions is
249 unchanging. In this way change might be described as occurring only in the relative
250 weighting of each component function (its coefficient in the functional basis we have
251 extracted from the data) needed to represent the trait. Since now we are only
252 studying the pattern of changes in a vector of weights, methods that model the

253 evolution of univariate and multivariate characters [6-9] can be used rather than the
254 full Gaussian process framework described in the boxes. We note that these
255 approaches still require a choice of covariance structure between points in time.
256 Again these methods can help infer phylogenies and/or generate hypotheses about
257 ancestral function-valued traits.

258

259 **Evolutionary inference for speech-sound data**

260

261 With the aim of performing function-valued evolutionary inference with speech-sound
262 data the best functional representation of these data remains open: we suggest
263 three possibilities. Considering intonation contours as evolving function-valued traits
264 might shed light on the relationships between tonal and non-tonal versions of speech
265 sounds. Spectrograms, which show how the different frequency components present
266 in the signal vary through time, are useful (2D) functional representations of speech
267 sounds since these allow approximate reconstruction of the speech sound from their
268 forms; a similar case holds for coefficients derived from Linear Prediction Coding as
269 mentioned above.

270

271 Possible inhomogeneities in speech-sound evolution include abrupt changes, and
272 punctuated bursts of evolution [2]: we note that approaches to these possibilities
273 have already been considered in the context of single-character and multivariate
274 phylogenetic models of evolution [35]. Such approaches correspond, in our context,
275 to allowing the covariances between points in time to themselves be time-varying.
276 Although we regard these proposed approaches as crude (as first models should
277 be), a standard approach to compensate for a simplistic model is to consider traits

278 which are clearly homologous as this allows one to partly control for more exotic
279 forms of evolution not (presently) included in the model.

280

281 The evolution of speech may be considered at several different levels of resolution:
282 phonemes, words, or word groups. Historical linguists often consider the co-evolution
283 of groups of words which have phonetic similarities. In a Gaussian process
284 framework, co-evolution could be modelled by specifying appropriate covariances
285 between the words. By incorporating more information such joint approaches offer
286 the prospect of facilitating explicit hypothesis tests concerning the character of
287 individual ancestral spoken words.

288

289 **Directions and Data**

290

291 Now that we have made the case for a function-based treatment of data such as
292 speech and sound change, and noted ways in which existing methods and possible
293 variants could be used to make data-driven models for spoken language evolution, it
294 is appropriate to discuss more general questions of speech sound evolution data
295 and method. There currently exist several relatively uncontroversial linguistic
296 phylogenies, a natural starting point being the Indo-European languages [37,38,39].
297 These can be used for a first approximation to the evolutionary history of individual
298 spoken words. Contemporary samples of speech from languages at the most recent
299 branches of the linguistic evolutionary tree are readily available, e.g. [40].

300 Though it is not required for our approach, recordings of experts in ancient

301 languages - even though they are not native speakers, of course - could yield root

302 speech recordings useful for testing proximate hypotheses of speech-sound
303 evolution. Similar (homologous) words across languages have already been found
304 for a large number of words, by centuries of philological scholarship (e.g., [39]).
305 Existing methods can model (and so accommodate) variation in the duration of
306 recorded speech sounds through the use of warping and other registration
307 algorithms (e.g. [41,23]). From the tree and from tip-word recordings, it could thus be
308 feasible to infer and hence synthesize the sounds of words from any desired
309 generation, along any branch, through a statistical reconstruction of traditional
310 philological methods.

311 When analyzing empirical data, it is possible first to pool a training dataset to identify
312 the nature of variation. Random factors (e.g., in linguistics, different speakers), can
313 be taken into account at this stage using methods akin to those commonly used in
314 function valued trait analysis [34]. where explicit variational types (e.g., genetic and
315 environmental variability in function valued trait models) are modelled separately,
316 then combined. Variability may also be accounted for via mixed effects models, an
317 extension of linear models already considered for function-valued data in linguistics
318 [32]. Traits would then be subject to an evolutionary process that has been explicitly
319 adjusted to the patterns already observed in empirical data.

320

321 We have discussed how function-valued data might be used to perform both
322 phylogenetic inference and also to generate hypotheses about ancestral linguistic
323 traits. We might conclude by answering the question: why would an ability to
324 reconstruct past speech sounds be worthwhile? We believe that, as well as
325 connecting us with our past, moderate success in this endeavour could have a
326 substantial effect on historical linguistics. More broadly though, it is our contention

327 that these data-analysis approaches can make a positive contribution to the
328 investigation of evolutionary processes in a large number of biological contexts.
329 Whereas the treatment of continuous variables in the context of a phylogenetic
330 analysis was once considered unacceptable [42,43,44], there is a new interest not
331 only in using phylogenies to understand and test hypotheses involving the historical
332 development of continuous traits (e.g, via the comparative method [8,45,46]), but
333 also to allow continuous variables to participate directly in the inference of
334 phylogenetic relationships among species. [47,48,49]. We see our efforts to create
335 statistical tools that can be used to incorporate function-valued data into
336 phylogenetic data-analysis contexts as closely related to this (now) well-established
337 research programme. These methods hold the promise of extending the power of
338 phylogenetic analysis into new research fields, bringing new data to bear on the
339 many phylogenetic problems that have thus far resisted resolution under the
340 phylogenetic systematics paradigm, and providing new ways of treating old data to
341 maximize their contribution to the understanding of nature and of natural processes,
342 including those traits that have a critical role in shaping the evolutionary history of
343 our own species.

344

345 **7. References**

346 [1] Meyer, K. and Kirkpatrick, M (2005) Up hill, down dale: quantitative genetics of
347 curvaceous traits. *Phil. Trans. R. Soc. B* 360, 1443-1455

348 [2] Gray, R. and Atkinson, Q.D. (2003) Language-tree divergence times support the
349 Anatolian theory of Indo-European origin. *Nature* 426, 435-439

350 [3] Atkinson, Q.D. et al. (2008) Languages Evolve in Punctuational Bursts. *Science*

351 319, 588

352 [4] de Boer, B. (2000) [Self organization in vowel systems](#). J. Phonetics 28, 441–465

353 [5] Jones, N.S. and Moriarty, J.M. (2010) Evolutionary Inference for Functional Data:

354 Using Gaussian Processes on Phylogenies to Study Shape Evolution.

355 [arXiv:1004.4668](#)

356 [6] Felsenstein, J. (1985) Phylogenies and the Comparative Method. American

357 Naturalist 125, 1-15

358 [7] Felsenstein, J. (1988) Phylogenies and quantitative characters. Ann. Rev.

359 Ecology and Systematics 19, 445-471

360 [8] Martins, E. P. and Hansen T.H. (1997) [Phylogenies and the Comparative Method: A](#)

361 [General Approach to Incorporating Phylogenetic Information into the Analysis of Interspecific](#)

362 [Data](#). American Naturalist 149, 646-667

363 [9] Rasmussen C.E. and Williams, C.K.I. (2006) [Gaussian processes for machine](#)

364 [learning](#). MIT Press

365 [10] Allen, W.S. (1953) *Phonetics in Ancient India*. Oxford University Press

366 [11] Allen, W.S. (1965) *Vox Latina: the pronunciation of Classical Latin*. Cambridge

367 University Press pp. 95-100

368 [12] Allen, W.S. (1968) *Vox Graeca: the pronunciation of Classical Greek*.

369 Cambridge University Press pp. 162-7,

370 [13] Halliday, M.A.K. (1981) The origin and early development of Chinese

371 phonological theory. In *Towards a History of Phonetics*. (Asher, R.E. and Henderson,

372 E.J.A., eds), pp 123-140, Edinburgh University Press

373 [14] Huang, X. et al. (2001) *Spoken Language Processing: a guide to theory,*

374 *algorithm, and system development*. Prentice Hall PTR

375 [15] Öhman, S.E.G. (1966) Coarticulation in VCV utterances: spectrographic

376 measurements. *J. Acoust. Soc. America* 39, 151-168

377 [16] Liberman, A.M. (1970) The grammars of speech and language. *Cognitive*
378 *Psychology* 1, 301-323

379 [17] Olive, J.P. et al. (1993) *Acoustics of American English Speech: A dynamic*
380 *approach*. Springer-Verlag New York Inc.

381 [19] Tyack, P. L. (1998) Acoustic communication under the sea. In *Animal Acoustic*
382 *Communication: Sound Analysis and Research Methods* (Hopp, S.L. et al. eds), pp
383 163-220, Springer

384 [20] Janik, V. M. (1999) Pitfalls in the categorization of behaviour: A comparison of
385 dolphin whistle classification methods. *Anim. Behav.* 57, 133–143

386 [21] Suzuki, R et al (2006) Information entropy of humpback whale songs. *J. Acoust.*
387 *Soc. Am.* 119, 1849-1866

388 [22] Yu, A. (2006) Sound Change: Phonetics. In K. Brown, ed. *Encyclopedia of*
389 *Language and Linguistics (Second Edition)*. Elsevier. 525-528

390 [23] Ramsay, J.O. and Silverman, B.W. (2005) *Functional Data Analysis*. Springer
391 New York

392 [24] Ramsay, J. O. and Silverman, B.W. (2002) *Applied Functional Data Analysis*.
393 Springer New York

394 [25] Ferraty, F. and Romain, Y. (2010) *The Oxford handbook of functional data*
395 *analysis*. Oxford University Press

396 [27] Griswold, C. et al (2008) Hypothesis Testing in comparative and experimental
397 studies of function-valued traits. *Evolution* 62, 1229–1242

398 [28] Gracco V. L. et al. (1996) Functional data analyses of lip motion. *J. Acoust. Soc.*
399 *Am.* 6, 3718-3727

400 [29] Koenig, L.L. et al. (2008) Speech production variability in fricatives of children

401 and adults: results of functional data analysis. *J. Acoust. Soc. Am.* 5, 3158-3170

402 [30] Gubian, M. et al. (2009) FDA as a Tool for Analyzing Speech Dynamics: A Case
403 Study on the French Word *c'était*. In *Proceedings of the 10th Annual Conference of*
404 *the International Speech Communication Association (Interspeech 2009)*, 2199-2202

405 [31] Grabe E. et al. (2007) Connecting Intonation Labels to Mathematical
406 Descriptions of Fundamental Frequency. *Language and Speech* 50, 281-310

407 [32] Aston, J.A.D. et al. (2010) Linguistic pitch analysis using functional principal
408 component mixed effect models. *J. Royal Stat. Soc. C* 59, :297-317

409 [33] Hadjipantelis, P.Z. et al. (submitted) Characterizing fundamental frequency in
410 Mandarin: A functional principal component approach utilizing mixed effect models.
411 CRiSM Research Reports, University of Warwick, 2011.

412 [34] Kirkpatrick, M. and Heckman, N. (1989). A quantitative genetic model for
413 growth, shape, reaction norms and other infinite-dimensional characters." *J Math*
414 *Biol*, 27: 429-450.

415 [35] Hansen T.F. and Martins E.P. (1996) Translating Between Microevolutionary
416 Process and Macroevolutionary Patterns: The Correlation Structure of Interspecific
417 Data. *Evolution* 50 (4) , 1404-1417.

418 [36] Fleming, W. H., Viot, M. (1979) Some measure-valued Markov processes in
419 population genetics theory. *Indiana University Mathematics Journal* 28 (5), 817–843

420 [37] Schleicher, A. (1861-2) *Compendium der vergleichenden Grammatik der*
421 *indogermanischen Sprachen*. Böhlau, Weimar

422 [38] Ringe, D. et al. (2002) Indo-european and computational cladistics. *Transactions*
423 *of the Philological Society* 100 (1), 59-129

424 [39] Mallory, J. P. and Adams, D. Q. (2006) *The Oxford Introduction to Proto-Indo-*
425 *European and the Proto-Indo-European World*. Oxford University Press

- 426 [40] Sample language corpora: http://www.audiofrench.com/vocabulary/vocab_index.htm,
427 <http://catalog.elra.info>
- 428 [41] Lucero, J., & Koenig, L. L. (2000). Time normalization of voice signals using
429 Functional Data Analysis. *J. Acoust. Soc. Am.* 108,1408-1420.
- 430 [42] Pimentel, R.A. and Riggins, R. (1987) The nature of cladistic data. *Cladistics* 3,
431 201-209.
- 432 [43] Felsenstein, J. (1988) Phylogenies and quantitative characters. *Annual Review*
433 *of Ecology and Systematics* 19, 445-471.
- 434 [44] Bookstein, F.L. (1994) Can biometrical shape be a homologous character? In
435 *Homology: The Hierarchical Basis of Comparative Biology* (Hall, B.K., ed), pp. 197–
436 227, Academic Press
- 437 [45] Harvey, P.H. and Pagel, M.D. (1991) *The comparative method in evolutionary*
438 *biology*, Oxford University Press.
- 439 [46] Rohlf, F.J. (2001) Comparative methods for the analysis of continuous variables:
440 geometric interpretations. *Evolution* 55 (11), 2143–2160.
- 441 [47] MacLeod, N. (2002) Phylogenetic signals in morphometric data. In *Morphology,*
442 *shape and phylogeny* (MacLeod, N. and Forey, P.L., eds), pp. 100–138, Taylor &
443 Francis.
- 444 [48] Catalano, S.A. *et al.* (2010) Phylogenetic morphometrics (I): the use of landmark
445 data in a phylogenetic framework. *Cladistics* 26, 539–549.
- 446 [49] Goloboff, P.A. and Catalano, S.A. (2011) Phylogenetic morphometrics (II):
447 algorithms for landmark optimization *Cladistics* 27 (1), 42–51

448

449

450 **Glossary:**

451 f_0 - **Fundamental frequency**. From an acoustic (sound) perspective, a speech
452 signal is a complex periodic wave composed of multiple sine waves. In voiced
453 sounds, the lowest frequency of this complex wave is its f_0 , the frequency at which
454 the vocal folds are opening and closing.

455

456 **(Empirical) covariance function (empirical covariance surface)** - The covariance
457 surface of a function is the functional analogue of a covariance matrix for multivariate
458 data. It must be positive or at least non-negative definite and in general is assumed
459 to have finite summable eigenvalues giving constraints on its nature (which must be
460 preserved in any evolutionary process). The empirical covariance surface is the
461 covariance surface estimated directly from multiple realisations of the underlying
462 stochastic process (the data).

463

464 **Undershoot:** In a movement (or a dynamical model of that movement), "undershoot"
465 means that the movement failed to meet the intended position target. For example,
466 one may make attempt to grasp an object (e.g. a fruit hanging in a tree), but fail to
467 grasp it. One may attempt to kick a moving ball, but be too slow to reach it. Or, in
468 speaking, one may intend to open the mouth for a certain vowel, but because of the
469 rapidity of speech fail to achieve the intended articulation.

470

471 **Overshoot:** In a movement (or a dynamical model of that movement), "overshoot"
472 means that the magnitude of movement was excessive, given the intended position
473 target. For example, one may reach for a cup, but accidentally knock it over instead.
474 Or an infant, intending to say "sock", may through inexperience utter "dock" instead,

475 because they have overestimated how much to move their tongue tip.

476

477 **(Bayesian) Nonparametric model:** an approach to (Bayesian) model selection and
478 adaptation in which the model size is allowed to grow with data size. In these
479 techniques, while individual variables are typically assumed to belong to parametric
480 distributions, the number of parameters is not fixed.

481

482 **Gaussian process (GP):** a stochastic process whose marginal distribution at each
483 point is Gaussian.

484