nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-019-0663-x

Weak biases emerging from vocal tract anatomy shape the repeated transmission of vowels

Dan Dediu[1,2,3,4*], Rick Janssen[2] and Scott R. Moisik[2,5]

Linguistic diversity is affected by multiple factors, but it is usually assumed that variation in the anatomy of our speech organs plays no explanatory role. Here we use realistic computer models of the human speech organs to test whether inter-individual and inter-group variation in the shape of the hard palate (the bony roof of the mouth) affects acoustics of speech sounds. Based on 107 midsagittal MRI scans of the hard palate of human participants, we modelled with high accuracy the articulation of a set of five cross-linguistically representative vowels by agents learning to produce speech sounds. We found that different hard palate shapes result in subtle differences in the acoustics and articulatory strategies of the produced vowels, and that these individual-level speech idiosyncrasies are amplified by the repeated transmission of language across generations. Therefore, we suggest that, besides culture and environment, quantitative biological variation can be amplified, also influencing language.

Language—similarly to other aspects of culture—is an evolutionary system in its own right, constantly shaped by adaptive pressures and neutral processes[1,2]. There are currently about 7,000 spoken languages[3], an essential aspect of this diversity being represented by their speech sounds (phonetics and phonology). There is wide cross-linguistic variation at this level[4], and a crucial question concerns the factors and processes driving the emergence and maintenance of this diversity[5]. Most sound changes are due to language-internal factors, such as co-articulation and misperception[6], but recent studies suggest that external factors might also generate pressures to which sound systems adapt[5]. As such, it has been suggested that aspects of the physical environment that vary spatially (e.g., altitude or air humidity) affect the physiology of speech production differently in different populations, resulting in differences between the speech sounds that occur in different languages[5,7,8].

However, our own cognitive, physiological and anatomical biases are probably the most important components shaping languages. Biases that are shared by all humans result in linguistic universals and universal tendencies[6,9]. However, in previous work we have argued that the extensive inter-individual variation that exists at all levels—from the molecular to the anatomical, physiological and neuro-cognitive—also plays a role in the emergence of cross-linguistic variation[5,10].

There is widespread variation at all levels between individuals and groups, including in genetics, anatomy and physiology, arising from our complex evolutionary history[11–15]. Here, we focus on variation in the morphology of the vocal tract (VT; see Fig. 1), which, despite the rather sparse evidence[5,16–19] is no exception[20–23]. Using high-quality data from a large multi-ethnic sample, we show that the oral part of the VT has overlapping but statistically distinguishable patterns of variation between participants from four broad ethno-linguistic groups. As we argue in detail elsewhere[16,17], variation in VT anatomy can produce articulatory biases that survive compensatory mechanisms, and that result in subtle acoustic or coarticulatory effects[19,24]. These weak effects can be amplified by the repeated use and transmission of language, influencing the processes of sound change and ultimately affecting the patterns of linguistic diversity[5,16]. However, this is an extremely complex, long and heterogeneous causal path with feedback loops, which must be investigated using methodologies and data from several scientific disciplines[10,16,25]. We have previously shown, using biomechanical modelling, that click production is affected by the shape of the alveolar ridge[17], and that the covert articulatory strategies used by non-native participants to produce the North American English 'r' sound is influenced by the shape of their hard palate[16].

We test here the hypothesis that the usually weak effects of such 'idiosyncratic' variation may be amplified through the repeated use and learning of language in groups where these variants are frequent enough, resulting in differences between the languages of groups with different biases[5]. While the cultural amplification of weak biases is supported by abstract modelling[26,27] and experiments involving universal properties of human cognition[28,29], we use here a more realistic model, where a detailed geometric simulation of the vocal tract built from actual human data is used to learn and produce vowels widely attested cross-linguistically. In this model we can precisely control the VT anatomy and observe (and propagate) its effects on the production of actual speech sounds. We concentrate here on one component of the oral VT, the hard palate (HP, which is under genetic and environmental controls, and shows inter-individual and inter-group variation[5,19,22,23]), and in particular, on the midsagittal hard palate shape (MSHPS; Fig. 1).

We use a model of MSHPS we developed previously[19] that allows us to accurately describe the midsagittal shape of any human hard palate (and to generate novel ones) using only four meaningful parameters (angle, fronting, concavity and weight) controlling a customized Bézier curve. With this model, we imported MSHPSs from 107 normal human participants into a simulation that learns to articulate, to a very high accuracy, a set of five 'seed' vowels (simultaneously cross-linguistically widespread and extreme in their coverage of the human vowel space; Fig. 2). It does so by discovering the affordances and constraints of its articulators (such as tongue, jaw and lips) and the nonlinear mapping between configurations of

[1]Laboratoire Dynamique Du Langage UMR5596, Université Lumière Lyon 2, Lyon, France. [2]Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands. [3]Collegium de Lyon, Institut d'Études Avancées, Lyon, France. [4]Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands. [5]Linguistics and Multilingual Studies, Nanyang Technological University, Singapore, Singapore. *e-mail: dan.dediu@univ-lyon2.fr

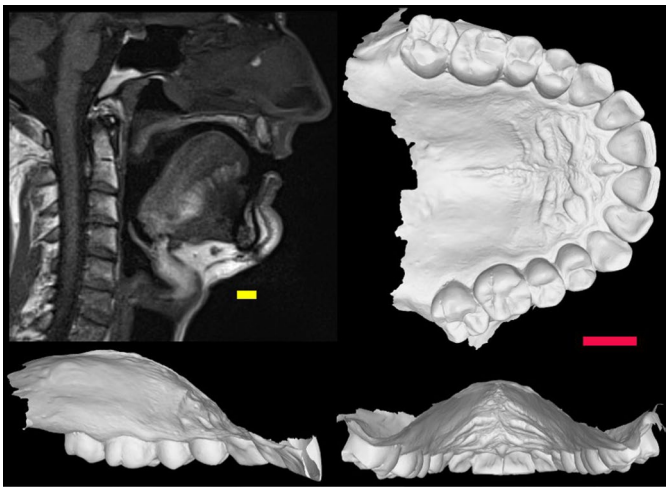

Fig. 1 | The shape of the human hard palate (the bony roof of the mouth). Top left: midsagittal MRI scan (yellow scale bar, 1 cm). The other three panels are projections of the three-dimensional (3D) intra-oral optical scan (red scale bar, 1 cm). Top right: transverse/inferior view (incisors to the right); bottom left: midsagittal/left view (incisors to the right); bottom right: coronal/posterior view (incisors away from viewer). Images not to scale (see the scale bars); the images are derived from the structural MRI and 3D optical intra-oral scans of author D.D., a 41-year-old male ($n=1$ participant). The top-left panel was created using Horos 2.4.0 (https://horosproject.org); the other three panels were created using MeshLab 2016.12 (http://www.meshlab.net), and the whole image assembled using GIMP 2.8.22 (https://www.gimp.org) on macOS 10.13 High Sierra.

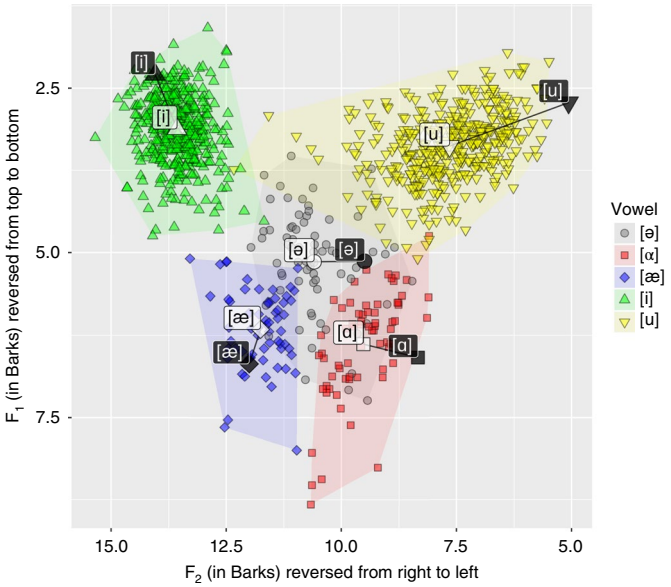

Fig. 2 | The distribution across languages and dialects of the five vowels used in the current study, highlighting the particular realizations used as 'seeds' for our models. The vowels are represented in the space of the first two formants $F_1 \times F_2$, where, as is customary for this type of plot: both axes are reversed, they show only the actual range of values excluding 0, and F_1 is represented on the *y* axis. The large black symbols (with associated white-on-black International Phonetic Alphabet (IPA) notations) are the canonical productions of the vowels we used as 'seeds' in our simulations ($n=5$ seed vowels): the mid central [ə] (as in American English 'uh' or 'sofa', with the values of the first five formants F_1– F_5 in Barks (5.13, 9.5, 14.56, 16.08, 18.04)); the high front [i] (as in 'beet' (2.29, 14.05, 15.63, 16.52, 17.88)); the low front [æ] ('bat' (6.69, 12.02, 14.69, 16.32, 18.9)); the high back rounded [u] ('boot' (2.72, 5.07, 15.14, 15.79, 16.96)); and the low back [ɑ] ('hot' (6.59, 8.34, 15.11, 15.91, 18.03)). The coloured symbols are actual realizations of these vowels recorded from several languages and dialects[31] ($n=1{,}051$ vowel realizations across 202 languages and 309 dialects); we also show their convex hulls (as transparent polygons) and means (large white symbols with associated black-on-white IPA notations). Our canonical vowels are more extreme than their cross-linguistic average realization (to which they are connected by black segments for easier visualization) and more extreme than even the most extreme actual realizations, as we want to start from extreme positions to maximally cover the potential human vowel space (especially for [i] and [u]). Moreover, actual productions during speech are influenced by multiple factors related to coarticulation, discourse and speed.

articulators and the produced sounds. This allowed us to precisely quantify the effects of observed variation in human MSHPS on how well vowels are learned, while keeping everything else constant. We focused on vowels because of the continuous (but cross-linguistically structured) nature of the human vowel space, the possibility to represent them by a small number of formants, and their relatively simple articulation. Variation in MSHPS produces biases that affect the articulation of vowels in subtle ways, resulting in weak but systematic inter-individual differences in the acoustics of the produced vowels. These effects are small within a generation (compared to the observed within-dialect and within-language variation in the realizations of these vowels). However, using the Iterated Learning Model paradigm[30] (where agents successively learn from the previous generation in a linear transmission chain), we show that these effects are amplified through cultural transmission. This results in differences between chains that are composed of agents with identical MSHPS within the chains, but different across them. Thus, the clustering of similar MSHPSs within groups, but different between groups, may lead to cross-linguistic differences between vowel systems.

Results

We present here the three main components of the results (see Methods and Supplementary Results 1 and 2 for methodological details and full results). First, we show that dense multidimensional measurements of oral VT anatomy do vary between broad ethno-linguistic groups, strongly supporting the fundamental assumption motivating these simulations. Second, we analyse a large cross-linguistic database of actual vowel productions, providing the proper background against which the results of our simulations should be understood. Finally, we focus on the simulations themselves, describing the quantification of the MSHPS using four-parameter Bézier curves. We show that our results are consistent between multiple independent runs for a select subsample of MSHPSs; we analyse the effects of MSHPS on the vowels produced at the end of the transmission chains and the amplification of the weak biases due to MSHPS by the repeated transmission along these chains; and we report on how the free articulators are used to partially compensate for variation in MSHPS.

We represent vowels by their first formants, and, while usually the first two (denoted F_1 and F_2) are used (given their importance for vowel perception), we also include the higher formants F_3, F_4 and even F_5, as fine details of the hard palate shape are more likely to affect higher resonances (associated with smaller wavelengths).

Oral VT morphology varies across groups. We analysed data on the anatomy of the anterior part of VT from more than 100 participants from four broad ethno-linguistic groups ('European' and 'North American of European descent', 'North Indian', 'South Indian' and 'Chinese'; these include the participants whose MSHPS

were used in the simulations). We found (see Supplementary Fig. 1, Supplementary Results 1 Part III and Supplementary Results 2) that the canonical variate analysis (CVA; a technique widely used for the analysis of multivariate data that computes linear combinations of the original variables that are not necessarily orthogonal but that maximize the ratio of between- and within-group variances) of 57 classical anthropological measurements has an overall classification accuracy of 84% for group and 94% for sex—whereas for the actual 3D intra-oral scans (IOS) of the lower jaw, CVA reaches 75% for group and 80% for sex, and for the upper jaw, 64% for group and 85% for sex. The Procrustes ANOVA with permutation (Supplementary Results 2, section 3.3.4.2) of the same IOS data for the lower jaw found that both main effects and their interaction are significant (Procrustes ANOVA with 1,000 permutations; for group: $F(3,86)=1.35$, 95% permutations confidence interval (95% pCI: the interval containing 95% of the F values obtained when permuting the data, that is, if the null hypothesis were true) is (0.79, 1.28), $P=0.007$, $Z=2.52$; sex: $F(1,86)=1.40$, 95% pCI (0.69, 1.48), $P=0.048$, $Z=1.82$; group × sex interaction: $F(3,86)=1.28$, 95% pCI (0.77, 1.30), $P=0.030$, $Z=1.96$), while for the upper jaw, only the main effects are significant (Procrustes ANOVA with 1,000 permutations; for group: $F(3,86)=1.46$, 95% pCI (0.77, 1.30), $P=0.005$, $Z=2.92$; sex: $F(1,86)=1.53$, 95% pCI (0.68, 1.53), $P=0.028$, $Z=2.09$; group × sex interaction: $F(3,86)=0.85$, 95% pCI (0.76, 1.32), $P=0.84$, $Z=-1.06$). Thus, the anatomy of the anterior part of VT shows differences between groups and sexes, but these differences are not sharp, are found only in highly multidimensional datasets, there are important overlaps between groups, and there are many 'mis-classified' individuals.

Real-world variation in the acoustics of the five vowels. We analysed a large database of vowel realizations (including [ə], [ɑ], [æ], [i] and [u]), containing mostly the first three formants F_1–F_3, across several languages and dialects[31] (Fig. 2 and Supplementary Results 1 Part I). We found that, as expected, there is important variation between the realizations of the 'same' vowel across languages (standard deviations (SD) between 0.33 and 1.01 with average 0.60 across all vowels and formants; all in Bark), but also within languages (average SD of 0.37, range 0.0–1.52) and dialects (average SD of 0.31, range 0.0–1.51). This variation is an important baseline against which to understand the results of our simulations.

Hard palate shape quantification for simulations. For our simulations, we modelled MSHPS using a Bézier curve with four meaningful parameters: angle, fronting, concavity and weight; more precisely, given a MSHPS tracing, we find the values of these parameters that produce a Bézier curve that best fits the MSHPS (detailed in ref. [19]). We conducted principal component analysis (PCA) on these four parameters across all 107 MSHPSs in our sample, and we found that the first two PCs (denoted as shape PCs and shortened to 'sPCs') explain 73.3% of the variance. Shape PC1 (44.9%) represents weight and fronting (contrasting higher, angled palates to shallower, smoother ones), and shape PC2 (28.4%) represents concavity and angle (contrasting flatter to more rounded shapes) (Supplementary Fig. 2). Using the methods described above, we found that these Bézier parameters do not capture the group and sex differences in our sample, which is probably due to their very low dimensionality, requiring much larger samples.

Consistency and variation across replications. Given that randomness plays a major role in many aspects of our simulations (especially in the learning mechanism), we checked the consistency of our results by re-running the whole transmission chain process 70 times for five selected MSHPSs: two artificial extremes created by manually setting the Bézier curve parameters to produce extremely low and extremely high (but still plausible) shapes, two actual MSHPSs from the ArtiVarK participants A87 and A73 (selected for their representativeness of the variation in that sample), and the average MSHPS across the participants in ref. [22] (Supplementary Fig. 3). This procedure ensures that we sample the possible evolutionary trajectories of the transmission chain for each of these five hard palate shapes. Reassuringly, we found that, in the final generation, the first five formants (F_1–F_5) of the tested vowels [ə], [ɑ], [æ], [i] and [u] have very narrow distributions across replications, with standard deviations between 0.02 and 0.43 (mean 0.10, median 0.05) Bark, and coefficients of variation between 0.001 and 0.08 (mean 0.01, median 0.004). Compared to the real-world within-dialect distances (Supplementary Results 1, section 2.2.1.4), the inter-replication differences for the actual MSHPSs (A87, average and A73) were significantly smaller (Tukey's post-hoc Honest Significant Difference tests across vowels and formants: $P<0.001$ for all three MSHPSs; see Supplementary Table 1), while for the two artificial extremes (low and high) were of a similar order of magnitude (see Fig. 3, Supplementary Fig. 4 and Supplementary Results 1, section 2.2.1). Moreover, the multidimensional scaling (MDS) and hierarchical clustering of the dynamic time warp distances (DWT) between all pairs of trajectories (that is, the time series of formant values over the generations of a chain) across MSHPSs, vowels and replications (Supplementary Results 1, section 2.2) found no systematic effects of replication. Therefore, we will focus here on the full dataset of 107 MSHPSs with a single replication each.

The effects of hard palate shape on the vowels in the final generation. The shape PCs affect the vowels in the final generation (Fig. 4 and Supplementary Fig. 5; see Table 1 for the linear regressions of individual formants on the shape PCs and their interactions for each vowel and formant): this influence is significant for [æ], [ɑ] and [i] for the first two formants, but covers all five vowels for the higher formants. The ethno-linguistic group, but not the sex, of the participant from which the MSHPS used in the model came makes a significant contribution to the final generation's acoustics, excluding F_1 and F_2 (see Supplementary Results 1, section 2.3.9). The F-tests comparing the regression models with and without group as predictor, separately for each formant, are as follows: F_1: $F(40,525)=1.15$, $P=0.246$; F_2: $F(40,525)=1.34$, $P=0.084$; F_3: $F(40,525)=2.16$, $P<0.001$; F_4: $F(40,525)=3.77$, $P<0.001$; F_5: $F(40,525)=2.93$, $P<0.001$. Likewise, considering the vowel system as a whole, group significantly affects the Procrustes distance to the 'seed' vowel system: $F(8,533)=6.15$, $P<0.001$. These effects are vowel-specific and concern especially [i] and [u].

The effects of hard palate shape are amplified across generations. The repeated transmission of language across generations results in the statistically significant increasing (but decelerating) divergence (measured by Procrustes distances; see Supplementary Fig. 6) of the five-vowel system as a whole from the original 'seed' system. These distances are very small (on the order of 0.14–0.24 interquartile range) when compared to a set of Procrustes distances between randomly generated five-vowel systems (ranging between 1.7 and 2.6), and their quadratic regression finds significant slopes β for both generation (linear regression slope $\beta=0.0065$, 95% CI (0.0059, 0.0071), $P<0.001$) and generation^2 ($\beta=-0.000078$, 95% CI (-0.000089, -0.000067), $P<0.001$), with an adjusted R^2 of 22% (Supplementary Results 1, section 2.3.6). Overall, the vowels become slightly more similar to each other, especially [u] (migrating towards [ə], [ɑ] and [æ] by almost 1.0 Bark) and less so for [i], while [ə], [ɑ] and [æ] are relatively stable; this quadratic trend is statistically significant but decelerating (Supplementary Fig. 7).

For a given vowel, formant and generation (denoted 'k'), we define drift_k as the formant value of the vowel produced at the end of generation k minus the seed formant value (Fig. 5 shows the

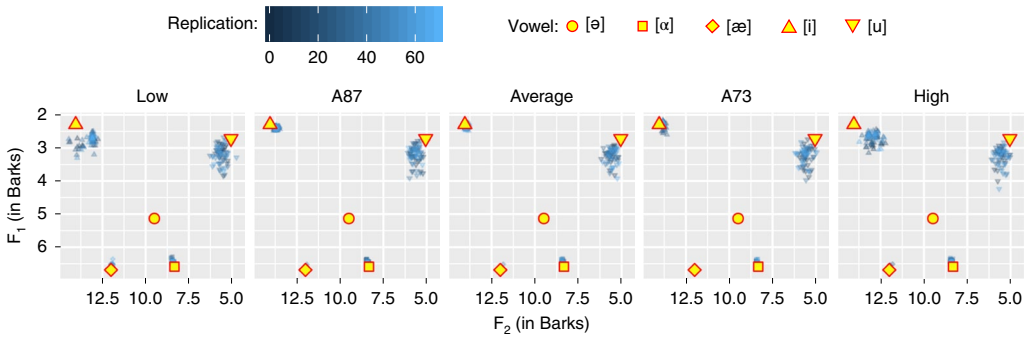

Fig. 3 | The distribution of the vowels in the final generation across replications for a selected set of five MSHPSs. Each panel shows the distribution (for one MSHPS named in the panel title) in the $F_1 \times F_2$ space (using the same conventions as in Fig. 2) of the seed (yellow symbols) and of actual productions in the final generation (blue symbols) across replications (the shade of blue). The dispersion across replications is small and varies among the vowels (the largest being for [u] and [i]) and MSHPSs (the two actual and the 'average' MSHPSs were capable of learning and transmitting our canonical [i]). Note that, for a given MSHPS, each replication represents a new run (thus using new pseudo-random values) of the same transmission chain composed of one agent per generation, for $n = 50$ generations, with all agents having the same MSHPS.

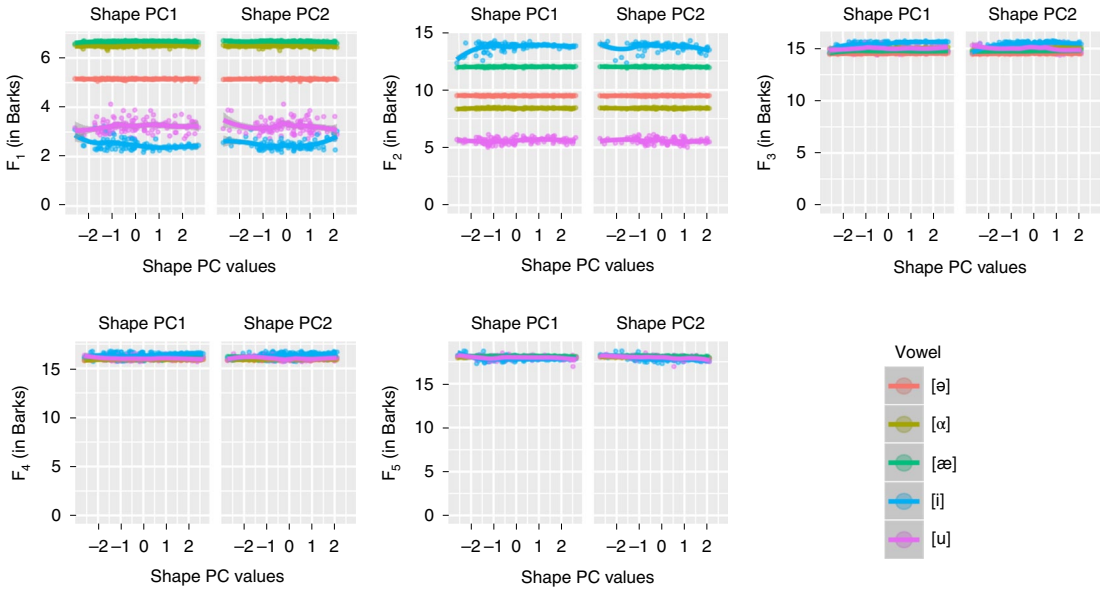

Fig. 4 | The dependency of the five formants (panels) on the shape of the hard palate (captured by the first two shape PCs; the twin columns in each panel) in the final generation for each vowel (colours). The *x* axis represents the values of the first two shape PCs describing the shape of the hard palate across participants, and the *y* axis the values of the first five formants; the coloured lines are the LOESS (locally estimated scatterplot smoothing) regressions (with 95% confidence intervals) per vowel. The vowel colour legend (bottom right) applies to all panels. Each dot represents the data for one of the $n = 107$ completed chains of 50 generations, where each generation comprises a single agent and all agents in a chain have the same MSHPS. See Supplementary Results 1 section 2.3.2.1 for a version of this plot where the scales and starting points of the *y* axes vary by formant to better capture the formant-specific range of variation (particularly important for the higher formants).

evolution of drift across generations). With this notation, the bias amplification due to repeated transmission is captured well by the slope of the linear regression of the drift after the whole chain was run ($drift_{50}$) on the drift after the first generation ($drift_1$). This is especially strong for [æ] and [i] (Supplementary Table 4), but applies generally: the linear regression across vowels, formants and MSHPSs has $R^2 = 15.8\%$, each generation amplifying the initial difference by the linear regression slope $\beta = 1.33$, 95% CI (1.21, 1.44), $P < 0.001$. Compared with the real-world vowel data for the first three formants, the within-dialect variation for the same vowel is significantly (as judged by the Bonferroni-corrected independent samples *t*-tests reported in Supplementary Table 4) larger than $drift_1$, but of comparable magnitude to $drift_{50}$ (which is, nevertheless, generally smaller than the within-language variation; see Supplementary Table 4). Moreover, $drift_{50}$ is significantly smaller than the inter-replication differences across vowels, formants and MSHSPs, and is positively correlated with $drift_1$ (mixed-effects models for the five replicated MSHPSs with $drift_1$, vowel, formant and their interactions as fixed effects and MSHPSs as a random effect with varying slopes for $drift_1$ gives a positive and significant fixed effect slope $\beta_{drift1} = 1.3$, 95% CI (0.61, 2.00), $P = 0.007$, and large inter-MSHPSs variation in slope with sd = 0.65; see Supplementary Results 1, section 2.2.1.5). Across MSHPSs, the final-generation vowels (except [ə]) are significantly less extreme than their seeds and closer to the actually observed real-world realizations, especially for [i] and [u] (compare Fig. 2, Fig. 3 and Supplementary Fig. 5; see Supplementary Table 4 for the Bonferroni-corrected independent samples *t*-tests comparing the 'seed', final and real-world

Table 1 | Linear regressions of the formant values of the final generation's vowels on quantitative measures of hard palate shape

Formant	Vowel	R^2	sPC1	sPC2	sPC1 × sPC2
F_1	[æ]	11.6%	0.003 (0.077)	**−0.008 (0.002)**	−0.002 (0.310)
F_1	[i]	12.9%	**−0.053 (<0.001)**	0.017 (0.358)	0.003 (0.860)
F_2	[ɑ]	5.0%	0.000 (0.957)	−0.004 (0.292)	**−0.007 (0.022)**
F_2	[i]	19.4%	**0.065 (0.003)**	**−0.065 (0.023)**	**−0.098 (<0.001)**
F_3	[ə]	12.7%	**0.007 (0.008)**	**0.009 (0.011)**	0.000 (0.935)
F_3	[ɑ]	61.2%	**0.012 (<0.001)**	**0.031 (<0.001)**	**−0.02 (<0.001)**
F_3	[æ]	20.2%	−0.001 (0.801)	0.005 (0.284)	**−0.018 (<0.001)**
F_3	[i]	34.5%	**0.085 (<0.001)**	**0.084 (<0.001)**	−0.023 (0.213)
F_4	[ɑ]	16.2%	−0.002 (0.374)	**0.007 (0.004)**	**−0.004 (0.043)**
F_4	[æ]	8.2%	0.000 (0.903)	**−0.011 (0.004)**	−0.005 (0.118)
F_4	[i]	17.9%	**0.058 (0.006)**	**0.105 (<0.001)**	0.018 (0.463)
F_5	[ə]	26.7%	−0.004 (0.223)	**−0.028 (<0.001)**	−0.002 (0.660)
F_5	[ɑ]	13.5%	**0.007 (0.011)**	**−0.010 (0.003)**	−0.003 (0.339)
F_5	[æ]	35.0%	−0.005 (0.279)	**−0.035 (<0.001)**	0.007 (0.184)
F_5	[i]	43.8%	−0.023 (0.068)	**−0.064 (<0.001)**	**0.092 (<0.001)**
F_5	[u]	19.1%	−0.022 (0.130)	**−0.077 (<0.001)**	0.018 (0.301)

Each row contains the regression of one formant for one given vowel on the first two shape PCs (and their interaction), showing the explained variance (R^2, in percent), and the βs (with uncorrected *P* values in parentheses) of the main effects (sPC1 and sPC2) and their interaction (sPC1 × sPC2). For economy reasons, the intercept α is not shown and we only include those rows (regressions) with at least one significant β (see Supplementary Results 1, section 2.3.2 for the full results); significant cells are in bold; the significance α-level is 0.05.

vowel realizations, and Supplementary Results 1, section 2.3.10 for more information and a different visual representation).

Articulatory compensation for hard palate shape. A PCA on the 11 articulatory parameters (of relevance here are jaw angle (JA), hyoid position (HX and HY), and positions of the tongue body, blade and tip (TCX, TCY, TBX, TBY, TTX and TTY); see Supplementary Methods and Supplementary Fig. 9 for details) across vowels and MSHPSs found that the first two PCs (denoted as articulatory PCs and shortened to 'aPCs') explain 50.6% of the variation and represent roughly the position and degree of tongue constriction: tongue height by aPC1 (28.1% of variation, with TCY, HX and TBY loading positively) and tongue fronting by aPC2 (22.6%, with TCX, TBX, HY loading positively, and JA negatively). These two articulatory PCs correlate strongly with acoustics, especially PC1 with F_1 (Pearson's $r(26748) = -0.905$, 95% CI (−0.907, −0.903), $P < 0.001$) and PC2 with F_2 (Pearson's $r(26748) = 0.857$, 95% CI (0.854, 0.860), $P < 0.001$), and change across generations in concordance with the changes in acoustics (Supplementary Fig. 8). However, the articulators also respond to variation in HP shape: their regression on the shape PCs (sPC1 and sPC2) while controlling for the acoustics in the final generation (Table 2) shows that it is mainly the tongue position being adjusted in response to HP shape.

Discussion

Using realistic computer models of the human vocal tract, we investigated if different midsagittal hard palate shapes (MSHPSs) resulted in different patterns of 'errors' in the production of these five vowels, and found that they did, but also that the effects were very small. Therefore, we investigated if the repeated transmission of vowels across generations (that is, their learning anew by naive agents using the productions of the previous generation as targets) would amplify these effects. We found that this amplification is a robust phenomenon, that it produces changes comparable to the observed patterns of vowel dispersion within real dialects and languages, and that it varies by MSHPS (producing final vowels that differ statistically across shapes and ethno-linguistic groups)—but also that the produced vowels 'move away' from the extreme canonical seeds we started with, towards realizations more typical of attested languages.

To check robustness, we ran 70 independent replications of the simulations for five selected MSHPSs, and found that the variation between replications is constrained around a central tendency (on the scale of real-world within-dialect variation), and that even artificially extreme MSHPSs produce effects within the range of the actually observed variation, but that there is a certain degree of contingency affecting individual replications. We represented the acoustics of vowels using the first five formants (thus going beyond the usual two or three), and we found that all vowels change across generations and for all MSHPSs following a decelerating trajectory, but that there are clear differences between the influence of MSHPS on different vowels and formants. Of the free articulators included in our vocal tract model, the most important ones were those controlling the position of the tongue body, blade and tip, being actively used to compensate for variation in MSHPS. However, this compensation is not perfect, and we found that the repeated transmission of the vowels across generations, far from dampening the weak influence of MSHPS, amplified it to an appreciable degree while producing overall more realistic vowels than the extreme initial exemplars we used to seed the transmission process.

Despite these results, our model described here has several shortcomings. First, the learning mechanism we used is neither ecologically valid nor very efficient, and while our decision is justified by our wish to use components that are as generic as possible to better isolate the sources of our results, alternatives such as intrinsic motivation and curiosity-driven learning[32] may be more appropriate. Second, our homogeneous, single-agent-per-generation chains are admittedly unrealistic and raise the possibility that the trans-generational amplification of weak biases we observed is probably too strong. However, our earlier abstract computational modelling[33,34] and theoretical[16] work suggests that far from dampening (or hiding) weak individual biases, complex settings (where whole communities of agents that have different anatomies use and transmit vowels across communicative networks over time) hugely complexify the preconditions and dynamics of their amplification. Moreover, our design here is purposefully simplified as much as possible, because

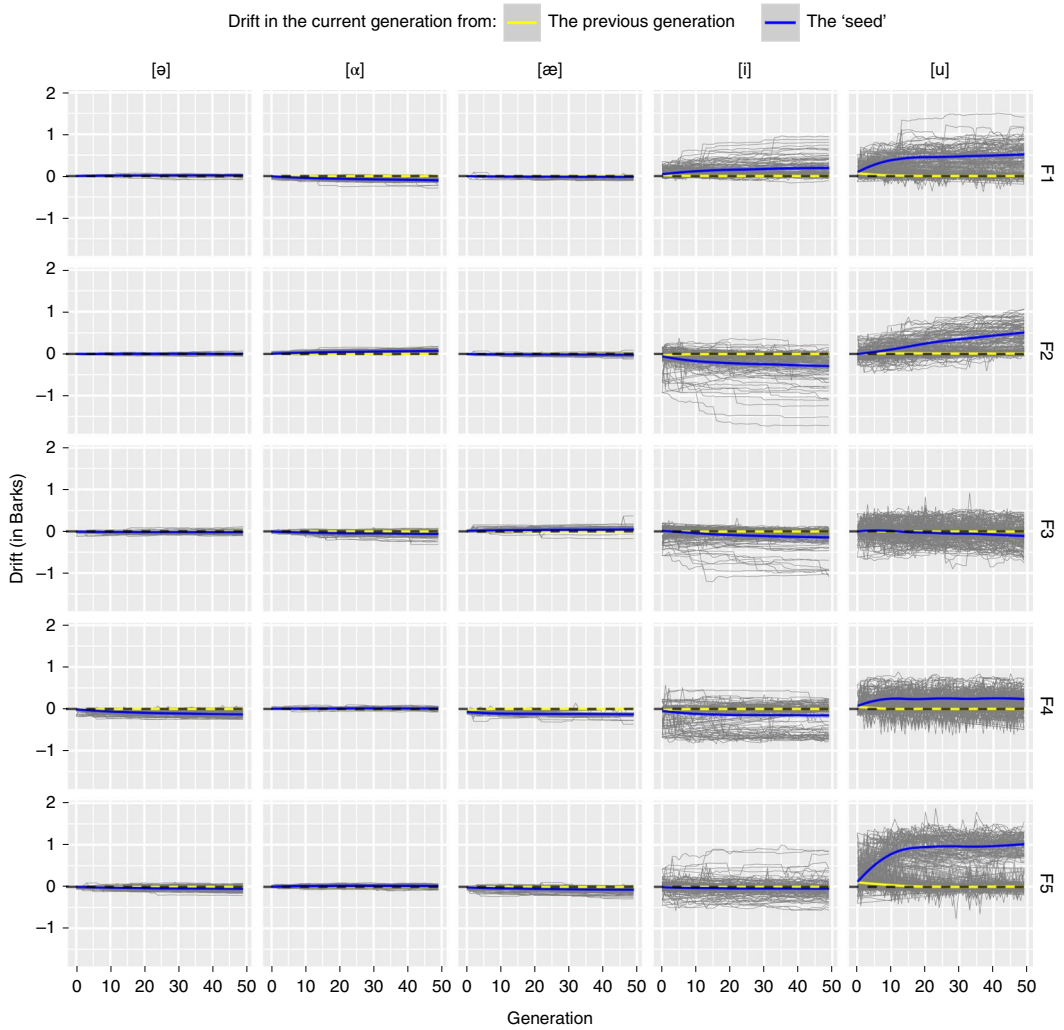

Fig. 5 | The amplification of weak biases through repeated transmission across generations. Each column shows one of the five vowels and each row one of the five formants, and each plot shows the evolution of drift across generations. In any given panel, each thin grey path represents the drift to the seed vowel in generation k ($drift_k$) in one of the $n = 107$ completed chains, while the thick blue curve is the LOESS smoothing across these paths. The yellow curve is the LOESS smoothing of the drift to the previous generation ($drift_1$) across all chains, and the black dotted line is 'no change' (that is, 0 drift). Each generation k comprises a single agent, and all agents in a chain have the same MSHPS. For different views, including magnification of the range of formant frequencies where the change takes place or focusing on the relationship between drift after the first generation and that after the whole chain was run, see Supplementary Results 1, section 2.3.7.

it is a necessary first step to ascertain that amplification works at all, before investigating it in realistic models with inter- and intra-group variation. Having established the necessary baseline, subsequent studies must test the deeper cultural evolutionary questions concerning intra-group variation. Third, we do not explicitly model phonology, co-articulation or language use in communication, focusing instead on individual vowels, but we do not expect such additions to invalidate the results reported here.

Taking our results at face value, how much of the observed linguistic diversity might be explained by such anatomical biases? On the one hand, there is extensive variation in hard palate shape between individuals that speak the same dialect, while, on the other, very dissimilar languages are spoken by groups with ostensibly similar distributions of hard palate shapes. Our own empirical data reported here show that there is continuous, multidimensional, overlapping variation between groups in the anatomy of the vocal tract. In our view, these phenomena are two sides of the same coin: as these biases are very weak, their effects are (in the vast majority of cases) successfully compensated and when not fully compensated, they are usually seen as speaker-specific idiosyncrasies

Table 2 | Articulators respond to hard palate shape

Vowel	Articulator	R^2	sPC1	sPC2
[ə]	aPC2	7.3%	0.013 (0.836)	**−0.195 (0.028)**
[ə]	TBX	6.7%	−0.002 (0.983)	**−0.267 (0.042)**
[ɑ]	TTY	13.0%	**−0.093 (0.007)**	**−0.119 (0.025)**
[æ]	TCX	39.6%	**−0.062 (0.001)**	**−0.063 (0.019)**
[i]	aPC1	19.5%	**−0.085 (0.036)**	−0.026 (0.645)
[i]	HY	17.8%	**0.029 (0.042)**	−0.004 (0.859)
[i]	TCX	28.2%	**0.207 (0.002)**	0.032 (0.725)
[i]	TCY	77.8%	**−0.079 (<0.001)**	**−0.113 (<0.001)**
[i]	TBY	64.5%	**−0.090 (<0.001)**	**−0.141 (<0.001)**
[u]	TTY	10.6%	0.016 (0.826)	**−0.285 (0.007)**

In the final generation, we regressed each articulator individually, as well as the two articulatory PCs (aPC1 and aPC2), on the first two shape PCs (sPC1 and sPC2), while controlling for the acoustics (the formant values), separately for each vowel. We show R^2 (in percent) and βs (P values in parentheses) only for those cases where at least one β is significant at $\alpha = 0.05$ (bold). See Supplementary Results 1, section 2.3.8.2 for full results.

(or speech defects), easily overridden by other forces of sound change. Nevertheless, in the right conditions (frequency of biased speakers, their positions in the communicative network and language-internal conditions, among others), the effects of such biases can be amplified and their effects phonologized[16].

The work presented here represents a first step in a long and inter-disciplinary research program[5,25]. After showing here (in a highly controlled and simplified setting) that vowels are sensitive to small differences in anatomy on the scale and pattern observed between individuals, the next steps should involve modelling of heterogeneous generations and experimental designs with human participants involving, for example, the reversible manipulation of hard palate shape[35]. The perspective advocated here argues that culture and biology co-evolve, and helps refocus interest from universalist and highly reductionistic explanations[36] towards an evolutionary view where variation is not noise, but a core feature of language and culture[1,37].

Methods

We implemented a computer agent that has a realistic model of the human vocal tract, where several articulators (for example, tongue tip or lip protrusion) can be moved to produce configurations resulting in actual speech sounds. The articulators are controlled by a neural network trained using a genetic algorithm to match as well as possible a set of predefined vowel targets (the 'seeds'). We can precisely control the MSHPS, and we investigated its impact on how accurately the seed vowels are learned and reproduced, within a single individual (ontogeny), and after the repeated transmission of language across generations (glossogeny).

Vocal tract model and hard palate shape. To model the shape of the human hard palate (the bony roof of the mouth; see Fig. 1), we modified the 3D geometric model of the vocal tract[38] VocalTractLab 2.1 (VTL2), which allows 11 articulatory parameters (Supplementary Fig. 11 and Supplementary Methods) to be manipulated and, for a set of parameter values, produces the corresponding sound. Internally, this is done by computing the area function corresponding to the shape of the vocal tract created by the position of the articulators by cutting this in planar sections along the airway centreline and computing the area of each section[38,39]. We added the capability to specify HP shape from actual human data (for example, MRI or intra-oral scanning) or programmatically, while maintaining the other aspects of the model unchanged. The MSHPS is described by a four-parameter Bézier curve model that we introduced previously[19] (Supplementary Methods and Supplementary Fig. 9), the transverse jaw curvature by a one-parameter cth-root curve, and the coronal profile by a one-parameter parabola, the last two being fixed here (Supplementary Methods and Supplementary Fig. 10[40]). Two MSHPSs were artificially generated extremes (one extremely 'low' and one extremely 'high', but still plausibly human) by setting the Bézier curve parameters manually. 107 MSHPSs resulted from fitting the Bézier curve model to tracings of actual human MRI structural scans using the procedure described in ref. [19]: 85 are participants in our own ArtiVarK study[18,19] (identified by 'A'; this is the subsample of participants with MRI scans used in ref. [19]: 32 female, age range 18–61, mean 27.2, median 24, SD 7.6), and 22 are from published MRI scans[22] (identified by 'T'; 10 female, age range 18–51, mean 29.8, median 27.5, SD 9.2). Finally, one MSHPS is fitted to the average of the 'T' MRI scans, denoted as 'average'. Supplementary Fig. 3 shows five selected MSHPSs (low, A87, average, A73 and high) with their midsagittal Bézier parameter values, and the fixed coronal and transverse profiles. The ArtiVarK study is covered by amendment 45659.091.14 (1 June 2015) "ArtiVarK: articulatory variation in speech and language" to the ethics approval "Imaging Human Cognition", Donders Center for Brain, Cognition and Behaviour, Nijmegen, approved by CMO Regio Arnhem-Nijmegen, The Netherlands (please see https://doi.org/10.5281/zenodo.1480426 for details). For Fig. 1, author D.D. has explicitly given his approval for the use of his MRI and intra-oral scans.

Learning to best approximate a set of given vowels. A vowel is described by its first five formants, $F_i \in R^+$, $i \in \{1,2,3,4,5\}$, measured on the psychoacoustical Bark scale[41]. For a given target vowel (which can be a predefined 'seed' vowel or a vowel produced by a different agent), described by its first five formants F_i^t, our goal is to train a naïve agent to produce a vowel F_i^p that best approximates the target, in the sense that it minimizes the Euclidean distance between them, $d_{tp} = ((F_1^t - F_1^p)^2 + (F_2^t - F_2^p)^2 + \ldots + (F_5^t - F_5^p)^2)^{1/2}$.

We implemented a feed-forward fully connected neural network with three layers, using the standard sigmoid activation function. The input layer has 6 neurons: 5 take the formants F_i^t of the target sound as input (scaled so that the approximate range of F_1– F_3, about 2 to 16 Bark, maps to the unit interval: $0.71 \times (F_i^t - 2) - 5$), the 6th neuron being a bias neuron. The hidden layer has 9 neurons, 8 receiving information from all 6 input neurons and the 9th being a bias neuron; all 9 are feeding information into the 11 neurons of the output layer. Each output neuron sets the value of one of the 11 free articulatory parameters of the VT model, which produces a sound with formants F_i^p. Thus, the neural network maps, through the synaptic weights w_{jk} of the connections between neurons, the first five formants of the target vowel, F_i^t, onto the formants of the produced sound, F_i^p.

We trained the neural network using a genetic algorithm approach, where a 'genome' is composed of 147 real-valued 'genes', each gene being one of the neural network's synaptic weights w_{jk}. The genetic algorithm minimizes the 'fitness' of the 'genome', defined as the Euclidean distance d_{tp} between the target vowel and the sound produced by the neural network with synaptic weights w_{jk} (or $+\infty$ if the articulatory configuration is impossible or no sound would be produced). We used a fixed population size of 100 'genomes' with stochastic universal sampling with elitism[42,43], and we used evolution strategies[44] to find a set of parameters controlling mutation. The genetic algorithm was run for a maximum of 500 iterations, unless the best fitness stabilizes across 100 successive iterations (an early stop). The genetic algorithm searches an enormous non-linear space defined by the 147 synaptic weights, attempting to find the neural network that maps the given target vowel to a configuration of the 11 free articulatory parameters of the VT that produces a sound closest to the target.

Seed vowels and language change across generations. We predefined five seed vowels: the mid central [ə] (as in American English 'uh' or 'sofa', with the values of the first five formants F_1–F_5 on the psychoacoustical Bark scale (which takes into account properties of human auditory perception) given by the vector of five numeric values (in Bark): (5.13, 9.5, 14.56, 16.08, 18.04)), the high front [i] (as in 'beet', (2.29, 14.05, 15.63, 16.52, 17.88)), the low front [æ] ('bat', (6.69, 12.02, 14.69, 16.32, 18.9)), the high back rounded [u] ('boot', (2.72, 5.07, 15.14, 15.79, 16.96)), and the low back [ɑ] ('hot', (6.59, 8.34, 15.11, 15.91, 18.03)), which are very frequent cross-linguistically[4] and better cover the phonetic space than the usual [i], [a] and [u] (Fig. 2).

For a given MSHPS c, we first trained a single naive agent to produce reproductions of the five seed vowels $v \in \{[ə], [i], [æ], [u], [ɑ]\}$ as well as possible. We denote this process as:

$$c: \{[ə], [i], [æ], [u], [ɑ]\} \rightarrow \{v^1_{[ə]}, v^1_{[i]}, v^1_{[æ]}, v^1_{[u]}, v^1_{[ɑ]}\}$$

where $v^1_{[ə]}$ is the actually produced sound for 'seed' vowel [ə], and so on. We further implemented an iterated learning Model (ILM)[30] whereby the vowels produced by the first generation of agents, $\{v^1_{[ə]}, v^1_{[i]}, v^1_{[æ]}, v^1_{[u]}, v^1_{[ɑ]}\}$, are used as targets for training a second generation of naive agents, whose productions are used to train a third generation, and so on until a predetermined final generation $n > 1$. These chains are homogeneous as the MSHPS of the agents, c, is conserved across generations:

$$c: \{[ə], [i], [æ], [u], [ɑ]\} \rightarrow \{v^1_{[ə]}, v^1_{[i]}, v^1_{[æ]}, v^1_{[u]}, v^1_{[ɑ]}\} \rightarrow$$
$$\{v^2_{[ə]}, v^2_{[i]}, v^2_{[æ]}, v^2_{[u]}, v^2_{[ɑ]}\} \rightarrow \ldots \rightarrow \{v^n_{[ə]}, v^n_{[i]}, v^n_{[æ]}, v^n_{[u]}, v^n_{[ɑ]}\}$$

We quantify the effects of MSHPS c and number of generations $1 \le k \le n$ by comparing the seed vowels $\{[ə], [i], [æ], [u], [ɑ]\}$ to the corresponding productions in that generation $\{v^k_{[ə]}, v^k_{[i]}, v^k_{[æ]}, v^k_{[u]}, v^k_{[ɑ]}\}$.

For any such chain, we extracted the trajectory of each of the formants for each of the vowels across generations (that is, the time series formed by the values $F_{v(i)}{}^k$ of the formant $1 \le i \le 5$ for vowel v in generation $1 \le k \le n$), and we computed the dynamic time warping distance between all corresponding pairs of trajectories (a lower distance captures similar trajectories). For a given vowel 'v', we computed the Euclidean distance (in the $F_1 \times F_2 \times F_3 \times F_4 \times F_5$ space) between its realization in generation $1 \le k \le n$, v^k, and the seed vowel v^0, and between v^k and its realization in the previous generation, v^{k-1} (for $k \ge 2$). For the vowel system as a whole (that is, the set of all five vowels) in generation k, we computed the ordinary Procrustes distances to the seed vowels and to the previous generation (for $k \ge 2$). Separately, we generated 100,000 random five-vowel systems in the $F_1 \times F_2 \times F_3 \times F_4 \times F_5$ space that respect the observed range of formant values, and we obtained a distribution of ordinary Procrustes distances between 0.24 and 4.51 (mean 2.18).

Because the learning procedure is computationally expensive, we ran a single replication (with 50 generations) for each of the 107 actual human MSHPSs, supplemented by 70 independent replications (each with 50 generations) for five MSHPSs selected for being extremes (low and high), representative for the observed human variation (A87 and A73) and the average of the observed human variation.

Within- and between-group variation in VT anatomy. We used three types of data: two derived from high-resolution 3D intra-oral scans (IOS) of 94 ArtiVarK participants (37 female, age range 18–61, mean 27.6, median 25, SD 8.1; note that not all participants with IOS also provided MRI scans due to, for example, orthodontic metallic implements), namely (1) the raw 3D coordinates of the IOS scans, and (2) a set of 57 'classical' anthropological measures derived from these IOS scans (angles, distances, ratios and regression coefficients; see Supplementary Results 1 section 3.1), as well as (3) the Bézier curve parameters[19] derived from the MRI scans of the 85 ArtiVarK participants and 22 North American participants from ref. [22] described above. The ArtiVarK study was designed with inter-individual and inter-group variation in mind, so that we aimed to recruit

about $n = 100$ participants in total, distributed as equally as possible between sexes and four broad ethno-linguistic groups ('European' and 'North American of European descent', 'North Indian', 'South Indian' and 'Chinese'), resulting in a convenience sample recruited through multiple channels (personal contacts, direct mailing and social networks) in the Netherlands and areas of Germany close to Nijmegen. Given the exploratory nature of the study, we did not conduct a formal power analysis, with the final sample size resulting from a balance between costs and access to MRI and intra-oral optical scanning facilities, and is larger than those usually reported in the literature[22–24]. As ArtiVarK is aimed at uncovering the patterning and effects of inter-individual and inter-group variation, we kept the conditions as uniform as possible across participants (that is, we did not have multiple experimental conditions and we did not randomize the participants beyond fluctuations due to recruitment). The VT anatomy data used in this paper was derived through a standardized procedure blinded to the aims of this study, as was the analysis. To test the presence of information about group and sex in these VT anatomical data, we applied both (1) exploratory methods (PCA and MDS) and (2) methods with a priori information about group and sex (MANOVA, random forests, CVA and Procrustes analysis). We found that, while the exploratory techniques do not recover the participant's group and sex from any data type, the methods with a priori information generally successfully recover significant differences between the groups using IOS data. The Bézier parameters have low dimensionality and cover only a specific aspect of VT anatomy, probably requiring a much larger sample than we currently have: a power analysis using G*Power 3.1[45] of the MANOVA test using the actual effect size with power $1-\beta = 80\%$ and significance level $\alpha = 0.05$, suggests a sample size of ≥ 27 participants per group (compared to our 10 'Chinese', 50 'Caucasian', 15 'North Indian' and 19 'South Indian' participants), and close to 300 per sex (compared to our 37 female and 57 male participants).

Distribution of [ə], [i], [æ], [u] and [ɑ] across languages and dialects. Building on shared code (https://github.com/soskuthy/u-fronting), we extracted up to the 4th formant for 1,051 actual realizations of the five vowels (72 for [ə], 438 [i], 63 [æ], 410 [u] and 68 [ɑ]) across 202 unique languages and 309 dialects from a published dataset[31] (see Fig. 2 and Supplementary Results 1 Part I), and we compared our simulated results to these data both in terms of spread for the same vowel and their central tendencies.

Computational details. For the results reported here, we used several high-end desktop computers (Intel Core i7-4790k 32 Gb RAM and Intel Core i7-3770 16 Gb RAM) and dedicated server blades (dual Intel Xeon E5-2620 64 Gb RAM), resulting in a maximum of 36 parallel execution threads, each thread implementing the learning of a single vowel for a given MSHPS; with this setup, more than 120,000 computer-days (more than 6 wall-clock months) were required.

Statistical analysis. All analyses are included and described in the reproducible Rmarkdown scripts, and all the results are contained in the HTML reports that result from the scripts' compilation, (see Supplementary Results 1 and 2). If not otherwise specified, we report uncorrected *P* values and the percent of adjusted explained variance R^2, and used two-tailed tests and an α-level of 0.05 for all statistical tests. We included all participants with the relevant data (MSHPS tracings derived from MRI scans, intra-oral scans and classical measurements, respectively; some ArtiVarK participants did not have MRI data due to issues such as orthodontic devices, but did have intra-oral scans, while none of the participants from ref. [22] have intra-oral scans or classical measurements). We conducted PCA across the four Bézier parameters (resulting in the 'shape PCs') and separately on the 11 free articulatory parameters (resulting in the 'articulatory PCs'). For the final generation of the chains, we performed linear regressions of the formant values on the 'shape PCs' and their interactions, and of the articulatory parameters (and the 'articulatory PCs') on the 'shape PCs' while controlling for the formant values, separately for each vowel (as the vowels show different chain dynamics). We performed multidimensional scaling (MDS) and hierarchical agglomerative clustering on the pairwise Dynamic Time Warping (DTW) distances between trajectories, and we conducted linear quadratic regressions of the Euclidean and Procrustes distances on the chain generation (these are supported by Generalized Additive Models (GAMs) shown in the Supplementary Results 1 Sections 2.2.5. "Generalized Additive Models (GAMs)" and 2.3.3. "The formants across generations"). Our data are relatively normally distributed (see Supplementary Results 1 Appendix V: Normality checks), with the notable exception of the Bézier parameters.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

All data, including the participant vocal tract anatomies, the MSHPS parameters and seed vowels (except for the 3D intra-oral scans, which are not provided because they may endanger our participants' privacy), are available in the Supplementary Information and in the GitHub repository https://github.com/ddediu/hard-palate-vowels.

Code availability

All the computer code of the simulations, the Rmarkdown scripts implementing the statistical analyses and plots, and a detailed 'How-To', are freely available in the Supplementary Information (Supplementary Software) and in the GitHub repository https://github.com/ddediu/hard-palate-vowels. The only exception is the modified source code of VTL2, available upon request under a custom license modelled on the original VocalTractLab 2.1 license; for this, only the pre-compiled version is freely distributable. The simulation software is written in C++, Java and Python2 and runs under Microsoft Windows 7 (or later), while the statistical analyses are implemented in R (embedded in Rmarkdown) and should run on any platform supported by these (Windows, macOS and various versions of Linux and BSD).

Received: 5 July 2018; Accepted: 21 June 2019;
Published online: 19 August 2019

References

1. Dediu, D. et al. in *Cultural Evolution: Society, Technology, Language, and Religion* (eds. Richerson, P. J. & Christiansen, M. H.) 303–332 (MIT, 2013).
2. Bentz, C., Dediu, D., Verkerk, A. & Jäger, G. The evolution of language families is shaped by the environment beyond neutral drift. *Nat. Hum. Behav.* **2**, 816 (2018).
3. Hammarström, H., Bank, S., Forkel, R. & Haspelmath, M. *Glottolog 3.2* (Max Planck Institute for the Science of Human History, 2018).
4. *PHOIBLE 2.0* (eds. Moran, S. & McCloy, D.) (Max Planck Institute for Evolutionary Anthropology, 2019).
5. Dediu, D., Janssen, R. & Moisik, S. R. Language is not isolated from its wider environment: vocal tract influences on the evolution of speech and language. *Lang. Commun.* **54**, 9–20 (2017).
6. Yu, A. C. L. *Origins of Sound Change: Approaches to Phonologization* (Oxford Univ. Press, 2013).
7. Everett, C., Blasi, D. E. & Roberts, S. G. Climate, vocal folds, and tonal languages: connecting the physiological and geographic dots. *Proc. Natl Acad. Sci. USA* **112**, 201417413 (2015).
8. Everett, C. Languages in drier climates use fewer vowels. *Front. Psychol.* **8**, 1285 (2017).
9. Christiansen, M. H. & Chater, N. Language as shaped by the brain. *Behav. Brain Sci.* **31**, 489–508 (2008).
10. Dediu, D. & Ladd, D. R. Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, ASPM and Microcephalin. *Proc. Natl Acad. Sci. USA* **104**, 10944–10949 (2007).
11. Reich, D. *Who We Are and How We Got Here: Ancient DNA and the New Science of the Human Past* (Oxford Univ. Press, 2018).
12. Jobling, M. A., Hollox, E., Hurles, M., Kivisild, T. & Tyler-Smith, C. *Human Evolutionary Genetics* (Garland Science, 2013).
13. Dediu, D. *An Introduction to Genetics for Language Scientists: Current Concepts, Methods, and Findings* (Cambridge Univ. Press, 2015).
14. Betti, L., Balloux, F., Amos, W., Hanihara, T. & Manica, A. Distance from Africa, not climate, explains within-population phenotypic diversity in humans. *Proc. R. Soc. B Biol. Sci.* **276**, 809–814 (2009).
15. Cramon-Taubadel, Nvon & Lycett, S. J. Human cranial variation fits iterative founder effect model with African origin. *Am. J. Phys. Anthropol.* **136**, 108–113 (2008).
16. Dediu, D. & Moisik, S. R. Pushes and pulls from below: anatomical variation, articulation and sound change. *Glossa J. Gen. Linguist.* **4**, 7 (2019).
17. Moisik, S. R. & Dediu, D. Anatomical biasing and clicks: evidence from biomechanical modeling. *J. Lang. Evol.* **2**, 37–51 (2017).
18. Moisik, S. R. & Dediu, D. Does morphological variation influence click learning and production? Evidence from a phonetic learning and imaging study. in *The Handbook of Clicks* (ed. Sands, B.) (Brill, in the press).
19. Janssen, R., Moisik, S. R. & Dediu, D. Modelling human hard palate shape with Bézier curves. *PLOS One* **13**, e0191557 (2018).
20. Howells, W. W. Cranial variation in man: a study by multivariate analysis of patterns of difference among recent human populations. *Pap. Peabody Mus. Archaeol. Ethnol.* **67**, 1–259 (1973).
21. Bosman, A. M., Moisik, S. R., Dediu, D. & Waters-Rist, A. Talking heads: morphological variation in the human mandible over the last 500 years in the Netherlands. *HOMO - J. Comp. Hum. Biol.* **68**, 329–342 (2017).
22. Tiede, M. K., Boyce, S. E., Holland, C. K. & Choe, K. A. A new taxonomy of American English /r/ using MRI and ultrasound. *J. Acoust. Soc. Am.* **115**, 2633–2634 (2004).
23. Zhou, X. et al. A magnetic resonance imaging-based articulatory and acoustic study of "retroflex" and "bunched" American English /r/. *J. Acoust. Soc. Am.* **123**, 4466–4481 (2008).
24. Brunner, J., Fuchs, S. & Perrier, P. On the relationship between palate shape and articulatory behavior. *J. Acoust. Soc. Am.* **125**, 3936–3949 (2009).
25. Dediu, D. in *Dependencies in Language: On the Causal Ontology of Linguistics Systems* (ed. Enfield, N.) 39–53 (Language Science Press, 2017).

26. Kirby, S., Dowman, M. & Griffiths, T. L. Innateness and culture in the evolution of language. *Proc. Natl Acad. Sci. USA* **104**, 5241–5245 (2007).
27. Thompson, B., Kirby, S. & Smith, K. Culture shapes the evolution of cognition. *Proc. Natl Acad. Sci. USA* **113**, 4530–4535 (2016).
28. Kirby, S., Cornish, H. & Smith, K. Cumulative cultural evolution in the laboratory: an experimental approach to the origins of structure in human language. *Proc. Natl Acad. Sci. USA* **105**, 10681–10686 (2008).
29. Culbertson, J. & Kirby, S. Simplicity and specificity in language: domain-general biases have domain-specific effects. *Front. Psychol.* **6**, 1964 (2016).
30. Kirby, S., Griffiths, T. & Smith, K. Iterated learning and the evolution of language. *Curr. Opin. Neurobiol.* **28**, 108–114 (2014).
31. Becker-Kristal, R. *Acoustic Typology of Vowel Inventories and Dispersion Theory: Insights from a Large Cross-linguistic Corpus* (Univ. California, 2010).
32. Moulin-Frier, C., Nguyen, S. M. & Oudeyer, P.-Y. Self-organization of early vocal development in infants and machines: the role of intrinsic motivation. *Cogn. Sci.* **4**, 1006 (2014).
33. Dediu, D. The role of genetic biases in shaping language-genes correlations. *J. Theor. Biol.* **254**, 400–407 (2008).
34. Dediu, D. Genetic biasing through cultural transmission: do simple Bayesian models of language evolution generalize? *J. Theor. Biol.* **259**, 552–561 (2009).
35. Brunner, J., Fuchs, S. & Perrier, P. The influence of the palate shape on articulatory token-to-token variability. *ZAS Pap. Linguist.* **42**, 43–67 (2005).
36. Berwick, R. C. & Chomsky, N. Why only us: recent questions and answers. *J. Neurolinguist.* **43**, 166–177 (2017).
37. Evans, N. & Levinson, S. C. The myth of language universals: language diversity and its importance for cognitive science. *Behav. Brain Sci.* **32**, 429–492 (2009).
38. Birkholz, P. Modeling consonant-vowel coarticulation for articulatory speech synthesis. *PLoS One* **8**, e60603 (2013).
39. Birkholz, P. *3D-Artikulatorische Sprachsynthese* (Logos Verlag, 2005).
40. Janssen, R. *Let the Agents do the Talking: On the Influence of Vocal Tract Anatomy on Speech During Ontogeny and Glossogeny* (Radboud University/ Max Planck Institute for Psycholinguistics, 2018).
41. Traunmüller, H. Analytical expressions for the tonotopic sensory scale. *J. Acoust. Soc. Am.* **88**, 97–100 (1990).
42. Baker, J. E. in *Genetic Algorithms and Their Applications: Proceedings of the Second International Conference on Genetic Algorithms* 14–21 (1987).
43. Eiben, A. E. & Smith, J. E. *Introduction to Evolutionary Computing* (Springer, 2003).
44. Beyer, H.-G. & Schwefel, H.-P. Evolution strategies–a comprehensive introduction. *Nat. Comput.* **1**, 3–52 (2002).
45. Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. G*Power 3: a flexible statistical power analysis for the social, behavioral, and biomedical science. *Behav. Res. Methods* **39**, 175–191 (2007).

Acknowledgements

We thank P. Birkholz for access to VocalTactLab 2.1's source code; our ArtiVarK participants; D. Norris and P. Gaalman for using the Avanto MRI scanner; T. Maal, F. Delfos and C. Kreulen for access to and help with the TRIOS intra-oral scanner; C. Jaques for participant recruitment and management; S. Kooijman for assistance with ethics; and M. Soskuthy for providing the community with the Becker-Kristal vowel corpus and base R code for its use. This work was funded by the Netherlands Organisation for Scientific Research (NWO) VIDI grant 276-70-022 to D.D., who was supported during the writing of this paper by a European Institutes for Advanced Study (EURIAS) Fellowship (2017-2018) and an IDEXLyon (16-IDEX-0005) Fellowship grant (2018–2021). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

R.J. wrote the computer code; S.R.M. created the seed vowels; S.R.M. and R.J. acquired the MRI and 3D intra-oral scanning data; S.R.M performed the non-rigid registration, landmarking, classical measures estimation, and initial CVA; R.J. and D.D. ran the simulations; D.D. performed the statistical analyses and plotting; D.D. and R.J. wrote the paper; R.J., D.D. and S.R.M commented on the paper; D.D. and S.R.M. designed and supervised the research; D.D. acquired funding.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-019-0663-x.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to D.D.

Peer review information: Primary Handling Editor: Marike Schiffer

nature research

Corresponding author(s): Dan Dediu
dan.dediu@univ-lyon2.fr

Last updated by author(s): Jun 16, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	A Python 2 script was used to fit the Bézier parameters to the manually traced MRI midsagittal hard palate shape. The simulation consists of a software chain comprising sections written in C++, Java and Python2. All these are available in the supplementary materials, the associated GitHub repository, and in already published materials, all clearly referred in the paper and the supplementary information.
Data analysis	The Rmarkdown scripts are fully available in the SI and in the GitHub repository. The analysis was run mainly on an Ubuntu 18.04 machine with R 3.4.4 and R 3.5.0, and was checked and replicated on separate machines running macOS High Sierra, macOS Mojave, Ubuntu 16.04, and Windows 10 Professional with R 3.4.4 and R 3.5.0 using different BLAS/LAPACK implementations (standard R, Apple Accelerate Framework and OpenBLAS).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All primary data are freely and fully available with the paper (in the SI and the GitHub repository) or are already fully and freely available as part of prior publications (as mentioned in the paper and the SI), with the exception of the 3D intra-oral scanning data which could potentially contain personal and/or identifiable information about our participants (for these, only the analysis script and its results are made public).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Quantitative experimental primary data derived from MRI and intra-oral scans of the vocal tract, and computer simulations.
Research sample	We used two samples. One comes a from published source [Tiede, M. K., Boyce, S. E., Holland, C. K. & Choe, K. A. A new taxonomy of American English /r/ using MRI and ultrasound. J. Acoust. Soc. Am. 115, 2633–2634 (2004)] containing data from 22 adult North American participants (10 female; age range 18-51, mean 29.8, median 27.5, sd 9.2). The other sample comes from our own ArtiVarK project and contains 85 adult participants (32 female; age range 18-61, mean 27.2, median 24, sd 7.6) recruited from four broad ethno-linguistic groups (Europe and North American of European Descent [speakeing Indo-European languages, mostly Germanic]; North India [speakers of Indo-Aryan languages]; South India [speakers of Dravidian languages]; and "Chinese" [speaekrs of Sino-Tibetan languages]). These data should be representative of the normal variation among adults in vocal tract anatomy.
Sampling strategy	Our own sample (ArtiVarK) was designed to cover language backgrounds of general interest, with targeted recruiting of participants from these backgrounds. We used multiple methods for recruiting participants (Facebook groups, direct e-mailing, announcements) from the Netherlands and areas of Germany close to Nijmegen. Given the exploratory nature of this study, we did not conduct a formal power analysis, with the final sample size resulting from a balance between costs and access to MRI and intra-oral optical scanning facilities, but it is larger than those usually reported in the literature. The full details about this sample are described in the paper and in cited published materials.
Data collection	For this study, we used only the static MRI and intra-oral scans of the participants (full details about the procedure are given in the article itself and the cited publications). The scanning procedure was standardized, and while the researcher was not blinded to the overall goals of the data collection (the effect of normal variation in vocal tract anatomy on speech production and cross-linguistic diversity), at the time the data was collected we did not yet fully design this particular study reported here (thus, there was implicit blinding in what concerns the influence of midsagittal hard palate shape on vowel transmission across chains of simulated agents, as reported here).
Timing	Our data was collected between April and September 2015. The published data comes from a paper published in 2004.
Data exclusions	We used all the participants with the appropriate data (MRI or intra-oral scans) -- please note that some could not provide MRI data due to various issues such as orthodontic metallic implements (the entire ArtiVarK database, including the test and calibration runs, contains a total of 94 participants, of which 85 could be used in this study).
Non-participation	The general non-participation rate is hard to estimate given the types of recruitment used, but among those that contacted us with an interest in participating, it was very low and related to conflicting scheduling or to situations (such as metallic orthodontic implements) that prevented safe MRI scanning.
Randomization	Given the focus of the study, namely inter-individual variation, we kept everything as constant as possible across participants. Thus, the tasks were identical, in the same order, and in the same conditions across all participants.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	See above
Recruitment	Adverts on campus and online, e-mailing campaigns, direct requests, and contacting facebook groups. Self-selection based on interest in language (and, more generally, science) is a possibility, but given that we were interested in inter-individual normal variation in vocal tract anatomy this is probably not a major issue.
Ethics oversight	The ArtiVarK study is covered by amendment 45659.091.14 (1 June 2015) "ArtiVarK: articulatory variation in speech and language" to the ethics approval "Imaging Human Cognition", Donders Center for Brain, Cognition and Behaviour, Nijmegen, approved by CMO Regio Arnhem-Nijmegen, The Netherlands (please see doi:/10.5281/zenodo.1480426 for details).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Magnetic resonance imaging

Experimental design

Design type	Static structural MRI scan of the vocal tract
Design specifications	Single structural scan obtained per subject with 10 minute acquisition time.
Behavioral performance measures	For the structural MRI scan, no behavioral measures were used.

Acquisition

Imaging type(s)	Structural
Field strength	1.5T
Sequence & imaging parameters	T1 MPR NS PH8, TE=2.58ms, TR=2250ms, flip angle 15°, slice thickness 1mm, pixel spacing 1mm×1mm, FOV 256×256
Area of acquisition	The vocal tract

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	MRI processed using MATLAB (Versions 2018b). Midsagittal hard palate traces obtained using a custom graphical user interface in Matlab of our own creation.
Normalization	Not applicable.
Normalization template	Not applicable.
Noise and artifact removal	Not applicable.
Volume censoring	Not applicable.

Statistical modeling & inference

Model type and settings	None relevant here
Effect(s) tested	None relevant here

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)	None relevant here
Correction	None relevant here

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☒	☐ Multivariate modeling or predictive analysis