

A thirteenth-century theory of speech

J. S. Harvey,^{1,2} H. E. Smithson,¹ C. R. Siviour,³ G. E. M. Gasper,^{4,5} S. O. Sønnesyn,^{4,5}

T. C. B. McLeish,^{6,7} and D. M. Howard⁸

¹*Department of Experimental Psychology, University of Oxford,*

Anna Watts Building, Radcliffe Observatory Quarter, Oxford, OX2 6GG,

UK^a

^{2b}*Department of Engineering Science, University of Oxford,*

Oxford e-Research Centre, 7 Keble Road, Oxford, OX1 3QG,

UK

³*Department of Engineering Science, University of Oxford,*

Oxford e-Research Centre, 7 Keble Road, OX1 3QG, Oxford,

UK

⁴*Department of History, Durham University, 43 North Bailey, Durham, DH1 3EX,*

UK

^{5c}*Institute for Medieval and Early Modern Studies, Durham University*

⁶*Department of Physics, University of York, Heslington, York, YO10 5DD,*

UK

^{7d}*Centre for Medieval Studies, University of York*

⁸*Department of Electronic Engineering, Royal Holloway,*

University of London, Egham Hill, Egham, TW20 0EX,

UK

^b Also at

^c Also at

^d Also at

(Dated: 4 July 2019)

1 In this historical paper we examine a pioneering theory of speech production and
2 perception from the thirteenth century. Robert Grosseteste (c.1175—1253) was a
3 celebrated medieval thinker, who developed an impressive corpus of treatises on the
4 natural world. Here we look at his treatise on sound and phonetics, *De genera-*
5 *tione sonorum* [*On the Generation of Sounds*]. Through interdisciplinary analysis
6 of the text, we find a theory of vowel production and perception that is notably
7 mathematical, with a formulation of vowel space rooted in combinatorics. Specifi-
8 cally, Grosseteste constructs a categorical space comprising three fundamental types
9 of movements pertaining to the vocal apparatus: linear, circular, and dilational-
10 constrictional; these correspond to similarity transformations of translation, rotation,
11 and uniform scaling, respectively. That Grosseteste’s space is categorical, and low-
12 dimensional, is remarkable vis-a-vis current theories of phoneme perception. As well
13 as his description of vowel space, Grosseteste also sets out a hypothetical framework
14 of multisensory integration, uniting the production, perception, and representation in
15 writing of vowels with a set of geometric figures associated with ‘mental images’. This
16 has clear resonances with contemporary studies of motor facilitation during speech
17 perception and audiovisual speech. We additionally provide an experimental foray,
18 illustrating the coherence of mathematical and scientific thinking underpinning this
19 early theory.

^{a)}joshua.harvey@pmb.ox.ac.uk

20 I. INTRODUCTION

21 This paper explores and responds to a historical theory pertaining to the psychology and
22 physiology of speech. This theory was developed in the early thirteenth-century, but within
23 it may be found many of the same considerations as those of modern neuroscience—the na-
24 ture of mental representations, the relationship between those representations and external
25 stimuli, and correspondences between the sensory faculties. Examining this theory, from
26 such a contrasting intellectual context to our own, raises questions of the role of experimen-
27 tation, observation, and modelling, and what constitutes permissible evidence for supporting
28 or rejecting hypotheses.

29 Robert Grosseteste (c.1175–1253) was a celebrated medieval thinker, who, as well as
30 writing on philosophy and theology, developed an impressive corpus of treatises on the
31 natural world. Here, we analyze one of these treatises—his text on sound and phonetics:
32 *De generatione sonorum* [*On the Generation of Sounds*] (*DGS*). The *DGS* was probably
33 written in the first decade of the thirteenth century, several centuries before the apparent
34 ‘scientific revolution in Early Modern Europe. It was a formative period, however, for the
35 development of European scientific thought, during which the reception of Greek natural
36 philosophy, enabled by their transmission, translations, and commentary from Arabic and
37 Greek into Latin, prompted new conceptual frameworks for the consideration of natural
38 phenomena^{1–3}. For modern science, reading medieval works presents several significant
39 challenges, starting not least with that of editions and translations. This analysis of the
40 *DGS* has only been possible through interdisciplinary collaboration between science and

41 humanities scholars, resulting in the compilation of a new critical edition and translation of
42 the text⁴⁵.

43 Previous interdisciplinary research has already explored other scientific treatises written
44 by Grosseteste: the *De colore* [On Colour]⁶, *De iride* [On the Rainbow]^{7,8}, and *De luce* [On
45 Light]⁹. In the *De colore*, Grosseteste develops a pioneering application of mathematics to
46 psychology. Within the space of approximately 400 words, he claims that colour occupies
47 a continuous, three-dimensional space, contrary to the prevailing one-dimensional theory
48 of the time⁶. It is surprising to find this theory articulated six centuries before three-
49 colour printing techniques were established¹⁰ and trichomacy was formulated by Thomas
50 Young¹¹. In the *DGS*, the treatise we explore and respond to in this paper, Grosseteste
51 attempts a similarly mathematical, combinatorial abstraction for phonetics—specifically for
52 vowels—as he attempts for colour. Several features of how he goes about doing this are of
53 interest to the modern reader. Whereas Grosseteste’s colour space is explicitly continuous,
54 the vowel space described in the *DGS* is explicitly categorical. Underpinning his theory
55 is a multimodal framework identifying correspondences between the mental representation
56 of vowels, their physical production, their perception, and their external representation as
57 letter shapes. Within this framework, the correspondences between speech perception, letter
58 perception, and shape perception, have particular modern resonances in audiovisual speech
59 and involvement of the motor system during speech perception. In the second half of this
60 paper we present an experimental interpretation of the text, using artificial vowel synthesis
61 and psychophysics to test the claims of correspondence between abstract, geometric acoustic
62 chamber shapes and vowel perception.

63 Before presenting a detailed discussion of the *DGS*, a question that might first be ad-
64 dressed is why one ought to concern themselves with medieval science. Modern neuroscience
65 is already at an interdisciplinary juncture between psychology, physiology, biology and math-
66 ematics; why should matters be further complicated with the inclusion of medieval history
67 and Latin? An answer may be found in the sheer wealth of scientific theory and observation
68 that was amassed during this period, which largely remains untapped. The history of science
69 is highly non-linear, despite its frequently linear presentation, leaving worthwhile questions
70 and suggestions unresolved in every historical age¹². Psychological phenomena such as the
71 perception of speech are not new, and have been prompting rational discourse throughout
72 many historical and geographical cultures. By engaging with these theories today, we may
73 find unexpected agreement with, or perspectives that are strikingly different to, our own.
74 In either case, we stand to gain much from the exercise.

75 II. ROBERT GROSSETESTE'S *DE GENERATIONE SONORUM*

76 The *DGS* begins with a physical description of vibrational mechanics: a sounding body
77 is such that when struck, its smallest parts move away from, return towards, and overshoot
78 their natural places, with vibrations occurring as a result. This is to be expected from the
79 given title of the treatise. However, only a quarter of the way through the text there is a
80 change of focus, as Grosseteste presents a case study of a particular sounding body, that is,
81 the production of human speech:

82 And since there is no such movement continuously in beings that have a soul,
83 such movement cannot come from a vegetative soul, but from a sentient motive

84 force and in a voluntary movement, which by necessity is preceded by the making
85 of a mental image or by apprehension. Therefore, a sound formed by a primary
86 motive force in which there is an ability to form mental images is a voice.

87 The remainder of the treatise is an attempt to characterize those ‘mental images’ that
88 initiate the voice, and the relationships between mental representations of origination, the
89 physical gestures of the vocal tract, the acoustic qualities of vowels, and the movements
90 of the hand that draw out letters to represent speech sounds¹³. Immediately following on
91 from the above passage, Grosseteste demarcates the difference between an intelligible and
92 an unintelligible speech sound:

93 But the actualising shaping itself of the vocal instruments and the shaping of the
94 movement of breaths able to move the vocal instruments gives to a certain voice
95 its kind and perfection; to a certain other voice, however, such shaping does
96 not give perfection. The voice, therefore, to which the aforementioned shaping
97 gives outward appearance and perfection, will be [called] a lettered voice. And
98 the voice that is completed by a single shape will be a letter. The voice that is
99 completed by several shapes will be composed of letters.

100 Here, Grosseteste establishes a direct relationship between the shapes—or as they may
101 be understood, *figures*—of mental images, vocal tract shapes, and the movements of the
102 breath during speech. These three figures, when perfected, give rise to a ‘lettered voice’,
103 i.e. an acoustic output of intelligible speech. Grosseteste does not yet describe these figures
104 geometrically, though that will come in the next section of the treatise. It is interesting

105 to note the particular emphasis on the natures of certain voices due to the ‘actualising
106 shaping itself of the vocal instruments’; any voice is preceded by a mental image, but the
107 intelligibility of that voice additionally depends on the speaker’s ability to precisely execute
108 the required motor programs. Or, to further unpack this notion, the acquisition of speech
109 requires first the presence of mental representations for speech sounds (it is unclear whether
110 Grosseteste is of the opinion these are innate or acquired), and second the learning of distinct
111 motor programs encoding muscular coordinations for the production of these speech sounds.
112 While Grosseteste does not explicitly describe this in terms of language acquisition, and the
113 development from an imperfect to perfected voice, it is heavily implied when understood
114 in the broader medieval context of discussions on the liberal arts. The seven liberal arts—
115 and in this case the first art, that of grammar—provide a means whereby the fallen and
116 corruptible things of the world may be refined and perfected through study and practice.
117 In this case, the notion of a ‘perfect’ or completed voice is related to the art, and study, of
118 grammar, and the acquisition of vocal tract coordinations that give rise to a ‘lettered’ voice,
119 i.e. intelligible speech¹⁴.

120 In isolation, it may seem from this passage that Grosseteste understands that both di-
121aphragmatic breath control (‘shaping of the movement of breaths’) and muscular coordina-
122tion of articulators (‘shaping of the vocal instruments’) are required to produce intelligible
123speech sounds. However, he later makes clear that he is instead claiming a direct iden-
124tity between control of the vocal apparatus and the resultant movements of the (‘motive’)
125breath, and it is these motive breath shapings that determine the ‘outward appearance and
126perfection’ of a voice. Writing six hundred years before Fourier and modern notions of fre-

127 quency, resonance, and spectral analysis, this provided a sensible hypothesis for the causal
128 relationship between the shape of the vocal tract and the acoustic qualities of the generated
129 sound.

130 Grosseteste then moves beyond the production of speech (the shaping of the vocal in-
131 struments and motive breaths) and its perception (its outward appearance) to the visual
132 representation of speech in writing, and in doing so provides further discussion on the nature
133 of these fundamental geometric figures:

134 The voice's capacity for being written down, therefore, is nothing other than this
135 same shaping of the vocal instruments and of the breaths by which the letter is
136 generated internally. It may therefore be represented by a visible shape similar to
137 the shape of its generation. It is clear, moreover, that, since art imitates nature
138 and nature always acts in the best possible way, and art does similarly when not
139 in error; however, representation by exterior shapes assimilated to interior will
140 be better than [representation done] otherwise: to write is, according to the art
141 of grammar, to represent interior shapes by means of exterior shapes similar to
142 these same interior shapes.

143 Here Grosseteste is guided by two Aristotelian principles: first, that 'art imitates nature',
144 or mimesis; and second, that nature always acts in the best possible way. There is clear
145 indication of his reading Aristotle's *De anima* [On the Soul]¹⁵, although Grosseteste does
146 not reference Aristotle directly, as he does in some other scientific works¹⁶. These principles
147 motivate one of the most central and clearly articulated claims of the treatise: the capacity
148 for speech to be written lies in the visual representation of shapes similar to the geometric

149 figures (mental, gestural, and of the ‘motive breaths’) at play during speech production which
150 is summarized in Figure 1. This claim that ‘representation by exterior shapes assimilated to
151 interior will be better than otherwise’ is particularly interesting, and has strong resonances
152 with recently resurfacing theories of non-arbitrary representation, or ‘iconicity’^{17,18}.

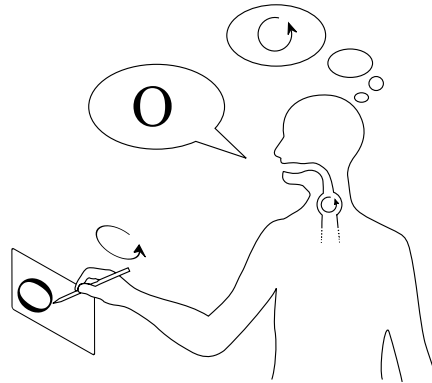


FIG. 1. A diagrammatic depiction of one of the claims in Grosseteste’s *De generatione sonorum*.

Grosseteste claims that the capacity for speech to be written lies in the visual representation of shapes similar to the geometric figures (mental, gestural, and of the ‘motive breaths’) at play during the production of speech. Because ‘art imitates nature’, the representational potential of letter shapes is maximized when those letters display geometric features common to the geometric figures at play in a vowel’s production.

153 For many languages today, including modern English, such a direct relationship between
154 speech-sound (phoneme) and written letter (grapheme) would be impossible; individual
155 letters have diverse pronunciations in differing lexical contexts, themselves quite different
156 to the letter name. As an example of phonological inconsistency, while an English speaker
157 with received pronunciation today may read the letter ‘O’ as a diphthong /əʊ/, it could be
158 similarly pronounced as /əʊ/ in ‘go’, but also as /u/ in ‘do’, /ʌ/ in ‘tonne’, /ʊ/ in ‘woman’

159 and even /I/ in ‘women’. This complication was not known to Grosseteste, who saw a mostly
160 direct and consistent grapheme-phoneme relationship in the languages it is likely that he
161 knew (Middle English, Latin, and French). Any exceptions, such as variations in regional
162 accents, could be accounted for as being ‘accidental’.

163 The treatise then gives a special consideration of vowels, for which Grosseteste provides
164 a comprehensive study of his hypothesized geometric figures.

165 The whole sound of the vowel and of any part of the vowel are the same as each
166 other. It is necessary, therefore, for it to be generated by a movement the parts
167 of which are the same as the whole. But there are seven movements in which
168 the parts are the same as the whole: straight movement, circular movement,
169 dilation and constriction—these last two do not differ except as straight move-
170 ment forwards and backwards—, circular movement over a centre in a straight
171 movement and a circular movement over a centre in a circular movement, and
172 likewise dilating and constricting movement over a centre in a straight movement
173 and over a centre in a circular movement.

174 In fact, this is a combinatorial system related to that described in the *De colore*: three
175 simple elements are combined in various ways to give rise to a full set including complex
176 combinations, except that for this scheme only two simple elements may be combined rather
177 than all three. It is also different in that, rather than being defined by independent dimen-
178 sions as in the case of the bipolar qualities of colours, only some of the simple elements
179 may be combined, and one—circular movement—may be self-combined. The choice of three
180 simple movements may not appear such an obvious choice, and it may be even more puzzling

181 why only one of the three may be self-combined. Grosseteste states clearly that this is the
 182 comprehensive list of movements ‘in which the parts are the same as the whole’. We may
 183 rephrase this description as one of time-invariant functions on position.

184 One way of interpreting the scheme that seems to resolve these confusions is by view-
 185 ing the three classes of simple movements as geometric linear transformations. In which
 186 case, these movements correspond perfectly to the allowed operations for Euclidean simi-
 187 larity transformations: straight movement for translation, circular movement for rotation,
 188 and dilational movement (and constrictional) as uniform scaling. Matrix notation provides a
 189 convenient and efficient way of describing these transformations; while Grosseteste would not
 190 have had this notation at his disposal, imagining these movements *per se* is not contingent
 191 on any particular form of mathematical description. Expressed as two-dimensional transfor-
 192 mation matrices of translation, rotation, and scaling— \mathbf{A}_t , \mathbf{A}_r , and \mathbf{A}_s , respectively—these
 193 three simple geometric transformations are given as:

$$\text{Translation : } \mathbf{A}_t = \begin{bmatrix} 1 & 0 & t \\ 0 & 1 & t \\ 0 & 0 & 1 \end{bmatrix} ; \text{ Rotation : } \mathbf{A}_r = \begin{bmatrix} \cos(t) & \sin(t) & 0 \\ -\sin(t) & \cos(t) & 0 \\ 0 & 0 & 1 \end{bmatrix} ;$$

$$\text{Scaling : } \mathbf{A}_s = \begin{bmatrix} t & 0 & 0 \\ 0 & t & 0 \\ 0 & 0 & 1 \end{bmatrix} .$$

194 Using this interpretive scheme, the geometric figures which Grosseteste describes natu-
195 rally arise by the consideration of points in Euclidean space experiencing these transforma-
196 tions. These simple and combined movements may be visualized in Figure 2 and Figure 3,
197 respectively, and in the videos included in the online version of this paper for translation
198 (Video 1), rotation (Video 2), dilation and constriction (Video 3), rotation and translation
199 (Video 4), and dilation/constriction and translation (Video 5).

200 [Mm. 1.](#) Translation. File of type mp4 (1.8 MB)

201 [Mm. 2.](#) Rotation. File of type mp4 (1.8 MB)

202 [Mm. 3.](#) Dilation and constriction. File of type mp4 (1.7 MB)

203 [Mm. 4.](#) Rotation and translation. File of type mp4 (1.8 MB)

204 [Mm. 5.](#) Dilation/constriction and translation. File of type mp4 (1.7 MB)

205 This interpretation also accounts for why straight movement does not give rise to a
206 distinct movement when self-combined, as the product of two translation transformations,
207 $\mathbf{A}_t^2 \mathbf{A}_t^1$, is simply another (different) translation, \mathbf{A}_t^3 . The same can be said for two consecu-
208 tive or simultaneous operations of scaling, or of dilational-constrictional movement. Circular
209 movements can, however, be self-combined to give a new class of self-similar movement, as
210 in Figure 4, and Video 6. The combination of circular movements over another circular
211 movement strongly connotes the epicyclic approach employed in classical and medieval as-

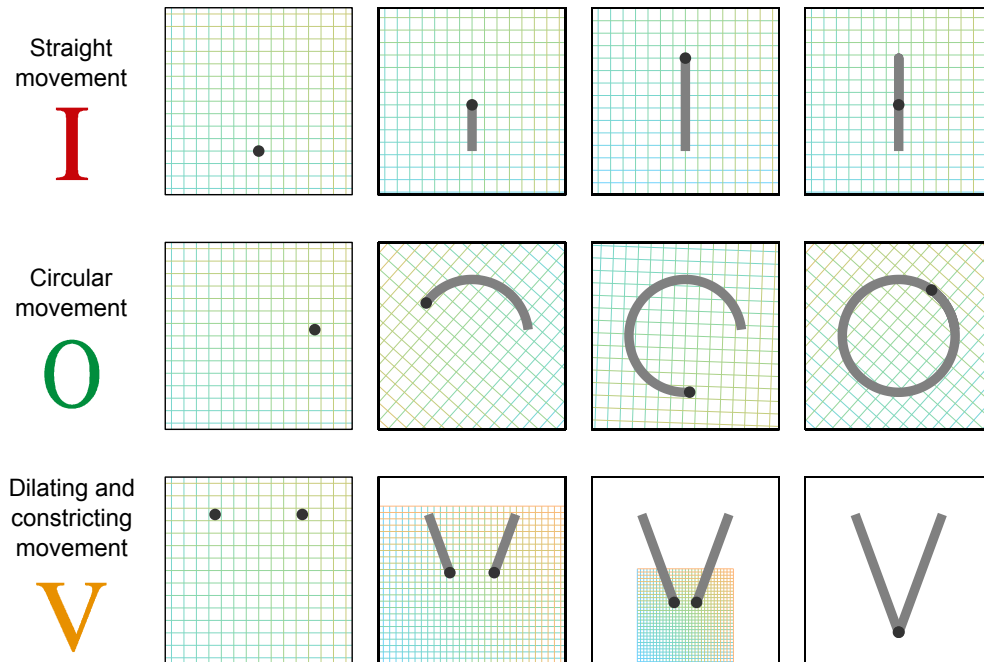


FIG. 2. The simple, self-similar geometric movements that Grosseteste describes as the basis for vowel categorization. We have interpreted his categories of simple movements—straight movement, circular movement, and dilating and constricting movement—as the three fundamental classes of linear geometric transformation: translation, rotation, and uniform scaling. Points (shown in black) embedded in planes undergoing these transformations trace out movements that agree well with Grosseteste’s descriptions of simple movements, shown in grey. Videos are provided in the online version of this paper.

212 tronomy, which comprises highly organized structures of rotating, nested spheres. In this
 213 case it is clear that an additional rotational transformation is applied to the space experi-
 214 encing the first rotational transformation, but the centre of this rotation is at a point offset
 215 from the origin, itself experiencing rotation. What first appears as an arbitrary selection
 216 of movements, in fact constitutes the complete scheme of self-similar, geometric similar-

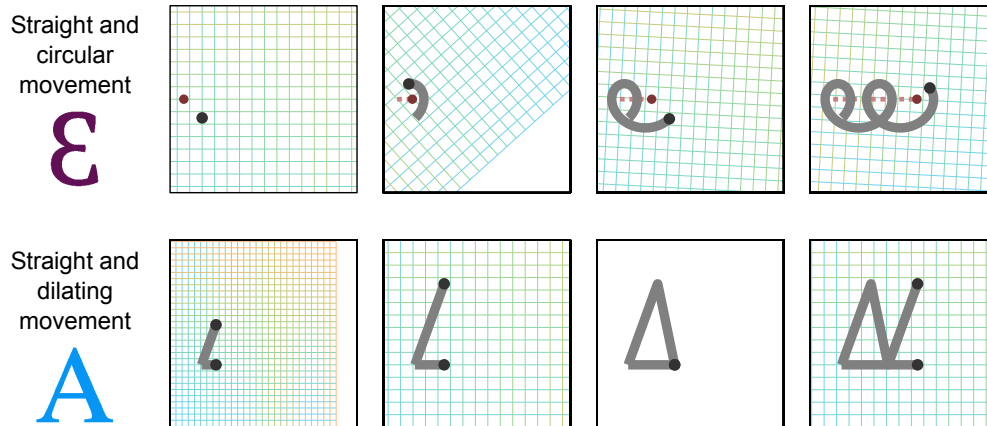


FIG. 3. The combined movements that give rise to vowels in Grosseteste’s model of phonetics. For the combination of straight and circular movement, the translating origin of rotation is indicated by a red dot. For the combination of straight movement with dilating and constricting movement, two dots repeatedly expand from, and collapse to, a single point that itself undergoes translation. Circular movement, or rotation, can be self-combined mathematically, as shown in Figure 4, but Grosseteste discounts it for vowel production as overly complex for the speaker. Videos are provided in the online version of this paper.

217 ity transformations of the two-dimensional plane, such that points in this plane trace out
 218 movements. However, to limit the number of vowels from seven to five (‘A’, ‘E’, ‘I’, ‘O’
 219 and ‘U’), Grosseteste discounts complex movements over a point itself tracing a circular
 220 movement—circular movements and dilational-constrictional movements over a centre al-
 221 ready experiencing circular movement are unfeasibly difficult:

222 [Mm. 6. Double rotation. File of type mp4 \(1.8 MB\)](#)

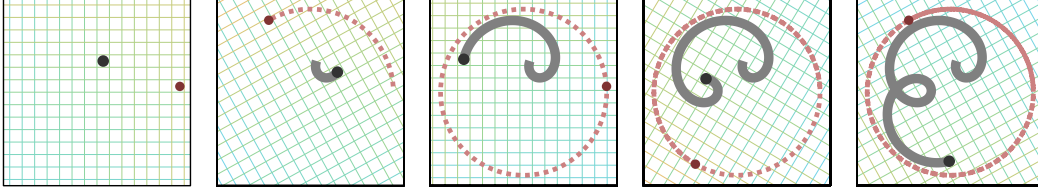


FIG. 4. Grosseteste describes a self-combination of circular movement, which he discounts as too complex for use in speech. This movement strongly evokes the mathematical constructions of epicycles in medieval astronomy. Here, the rotating origin of rotation is indicated with a red dot. Videos are provided in the online version of this paper.

223 On account of these seven movements the ancient Greeks posited seven vowels.
 224 But the abovementioned two movements over a centre in circular movement,
 225 granted that they are possible in imagination, are nevertheless difficult in reality.
 226 For this reason, there only remain five movements that are possible or easy to
 227 produce.

228 He then gives an in-depth geometric description of the remaining five self-similar move-
 229 ments, and how they generate the letters that represent their corresponding vowels:

230 It is therefore clear that in a straight movement of the motive breathings through
 231 the vocal tract an ‘I’ is shaped. But this straight movement is not a single contin-
 232 uous movement for then the lack of interruption would not cause a vibration but
 233 is very frequently coming and going. A circular movement over a centre makes
 234 the shape ‘O’. A circular movement over a centre [moved] in straight movement
 235 subtends a chord by the movement of the centre, and, by the movement of any
 236 point of the circumference, describes an arc over the chord and thus makes the

237 shape ‘E’. A constricting and dilating movement, on the other hand, makes the
238 figure ‘V’, that is, two lines running together in a centre. And a dilating and
239 constricting movement over a centre moved straight in a straight movement sub-
240 tends the base of a triangle. And any point, when there is dilation, because it is
241 moved by a double movement, describes one side of the triangle from the base
242 to the top, and when there is constriction, it describes the remaining side from
243 the top to the base, and thus it makes the figure ‘A’.

244 As shown in Figures 2 and 3, these descriptions align well with a linear transformation
245 interpretation of movement schemes. All five of the figures that Grosseteste traces out in
246 words can indeed be traced out by points or combinations of points embedded in the plane
247 experiencing the simple or combined similarity transformations of translation, rotation, and
248 or uniform scaling.

249 As made clear by these descriptions, the abstract figures that correspond to phonemes
250 (and, on account of the art of grammar imitating nature, graphemes) are not static geometric
251 shapes, but rather categories of movement, which are ascribed to the vocal tract during
252 speech. Therefore, for Grosseteste the perception of a speech sound, whether in hearing
253 speech or in reading, is intrinsically connected with vocal gestures, and the ‘mental images’
254 that encode their associated motor programs. This multisensory framework readily lends
255 itself to current discussions of the motor theory of speech¹⁹, and involvement of the motor
256 cortex in speech perception.

257 Eight centuries after Grosseteste was writing, we now have experimental evidence from
258 brain imaging and transcranial stimulation that his intuitions were solid. Involvement of

259 the motor system was established fifteen years ago in response to visual and auditory speech
260 perception²⁰, and soon after, that specific motor circuits in the precentral gyrus are recruited
261 to facilitate phoneme identification—serving as ‘speech-sound-specific neuronal substrates’
262 shared across the sensory and motor processes²¹. Motor cortex involvement has been found
263 to be beneficial for speech perception under noisy conditions²², and possibly under normal
264 listening conditions²³ (although possibly not²⁴). Of particular relevance to Grosseteste’s
265 theory, Möttönen and Watkins (2009) found direct evidence for motor representations play-
266 ing a complementary role in the categorization of speech sounds when they are found along
267 continua²⁵. As they point out, the mapping of highly variable acoustic signals onto discrete
268 motor representations could support the intelligibility of speech in challenging environments.
269 Even more intriguingly, Tian and Poeppel proposed a common sequential estimation mech-
270 anism underpinning both the quasi-perceptual experience of articulator movement and the
271 corresponding auditory percept of speech mental imagery²⁶. They claim that the experimen-
272 tal evidence from both task demands and stimulus properties demonstrates the top-down
273 role the motor system is playing in this type of mental imagery. In which case, Grosseteste’s
274 claim that the mental imagery of speech is in fact a mental representation not of sound, but
275 of motion (albeit of a simple, geometric nature), was remarkably apt.

276 In the light of these recent investigations, we can again consider Grosseteste’s approach
277 to understanding speech. Acoustic signals show enormous variety, and to the thirteenth-
278 century researcher writing before the advent of spectral analysis, this would have proved
279 impossible to organize. Confronted with the curse of dimensionality, Grosseteste limits his
280 study of sound to that of speech—a subset of natural sounds that the human auditory system

281 can reliably organize, doing so in a categorical manner. Aristotelian principles, the scientific
282 paradigm of the day, provide the methodological approach, with the movements of the hand
283 during writing perhaps constituting a permissible form of evidence for understanding the
284 mental and anatomical origins of speech, and its perception. That speech sounds differ due
285 to differences in movement category sits well with what Grosseteste understands about the
286 vibrational mechanics of sound; sound is the perception of a special class of movements made
287 by physical bodies, either when struck (the sounding body) or when formed by a primary
288 motive force capable of forming mental images (the voice).

289 **III. A PSYCHOPHYSICAL INTERPRETATION OF THE TEXT**

290 The claims in the *DGS* are bold, and may read today as ‘unscientific’, lacking any evi-
291 dential basis. But before dismissing these claims out of hand, it is worth considering exactly
292 what evidence would have been available at the time to a shrewd observer. The morphology
293 of the vocal tract would largely have been unknown, although from the end of the twelfth
294 century very good diagrams of the vocal tract and its articulators were being produced in
295 the Arabic-speaking world²⁷. These would not have been accessible to Grosseteste, and we
296 can reasonably say that any data he had regarding vocal tract morphology would have come
297 from his own direct experience of vision and proprioception. As has been remarked by oth-
298 ers, the resemblance of the ‘O’ letter shape and the pronounced rounding of the lips when
299 producing the /ɔ/ phoneme may suggest a non-arbitrary grapheme-phoneme relationship²⁸,
300 and could have been a motivating factor for the theory as a whole.

301 To experimentally determine whether Grosseteste’s theory could have been constructed
302 in a way commensurable with the available evidence, we created a set of synthetic vow-
303 els, using physical models of vocal tracts. These models were designed to incorporate the
304 geometric figures Grosseteste identified at the front of the mouth end of the tract. This
305 is, categorically, not to refute or accept the theory expounded in the *DGS*; we have ample
306 data on the morphology of the vocal tract, and nowhere does it feature idealized geometric
307 shapes as described in the *DGS*. However, in this manner we are able to evaluate whether
308 Grosseteste’s theory would have been consistent with the observational data available to
309 him—the visual and proprioceptive measurements of the mouth and lips. The question is,
310 therefore, not whether the theory is correct, but the following: can we construct acoustic
311 chambers that incorporate Grosseteste’s ideal geometric figures at the ‘mouth end’ (the end
312 furthest from the acoustic source), and yet are perceived as the five vowels in question?
313 We tested this using established methodologies of phonetics and speech perception, namely,
314 spectral analysis, and both multidimensional scaling and classification experiments.

315 **A. Stimuli**

316 Synthetic vowels were produced by plate-type model vocal tracts, constructed to resemble
317 the five geometric figures Grosseteste describes at the mouth end. This is a one-dimensional
318 model developed by Arai *et al.*²⁹, comprising 75 mm wide acrylic squares, each 10 mm thick,
319 with central holes of different diameters. The plates are clamped together in a specified order,
320 leaving a central cavity of varying size down the length of the tract. A rubber coupler allows
321 the introduction of an electrolarynx to acoustically stimulate the model at the laryngeal

322 end, which produces a falling pitch excitation in the male range from 100 Hz to 60 Hz
 323 lasting around two seconds. Adjustments were made to the laryngeal end of the models
 324 such that the output best approximated the associated phoneme. The resultant plates are
 325 shown in Figure 5, which also includes an overlay in red of the region made to resemble
 326 the geometric shape for each vowel. The acoustic outputs of these vocal tract models were
 327 then analyzed acoustically (formant analysis) and perceptually (two psychophysical listening
 328 tests), to evaluate how successfully the synthetic speech-sounds approximate natural vowels.

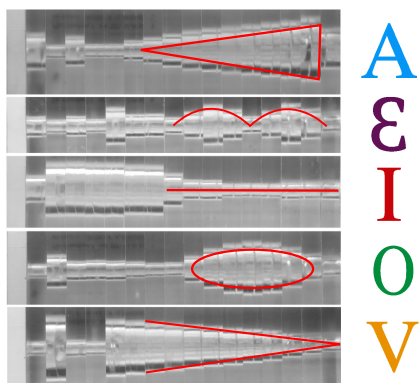


FIG. 5. The configurations of the plate-type vocal tract model (VMT-10) of Arai *et al.*²⁹ used to synthesize the five samples corresponding to Grosseteste’s geometric figure associations for each of the five vowel letters, with the mouth-end on the right. From top to bottom: A, E, I, O and V. The models are overlaid with the geometric shapes inferred from Grosseteste’s descriptions.

329 B. Formant analysis

330 Spectrograms for each sample were generated with a Hamming window of 20 ms, as shown
 331 in Figure 6, Upper Panel. The Lower Panel shows smoothed spectral slices calculated as

332 the mean of each spectrogram across time. Difference between these synthesized stimuli
 333 and natural vowels are the shape of the acrylic plates vs the speaker’s vocal tract—which is
 334 our primary interest—and the acoustic excitation (electrolarynx vs a speaker’s larynx). The
 335 electrolarynx for the Arai tubes provides a signal that has a constant spectrum whereas the
 336 output from the vibrating vocal folds of the speaker vary as a function of the airflow loading
 337 owing to the shape of the vowel being uttered, sub-glottal lung air pressure through breath
 338 control and the nature of the voice quality being employed and any pitch variation.

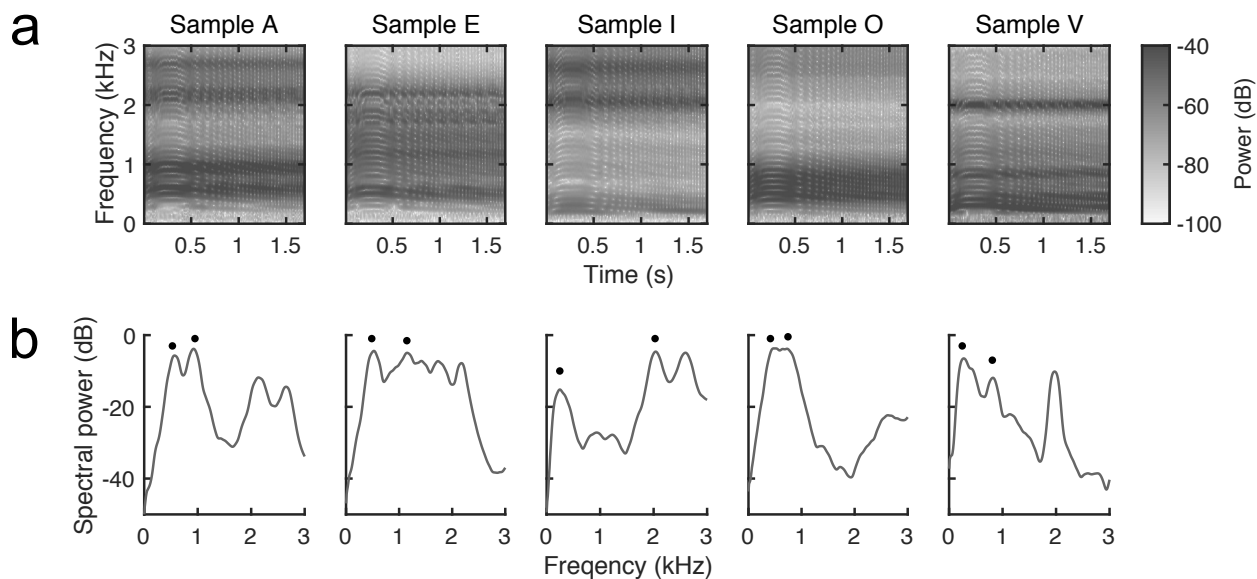


FIG. 6. (a) Spectrograms produced from each of the five synthesized samples. (b) Spectral slices given by the mean of each spectrogram across time for each sample, from which the frequencies of the first two formant peaks, F_1 and F_2 , were taken (indicated by black dots).

339 The horizontal dark bands in the spectrograms show formants (peaks in spectral power)
 340 that result from filtering the input acoustic excitation of the electrolarynx by the passive
 341 acoustic resonances of the chambers. The primary acoustic features of vowels are the lo-

342 cations in frequency space of their two lowest-frequency formants, F_1 and F_2 . When, for
343 different vowels, F_1 is plotted on the ordinate and F_2 is plotted on the abscissa, the vowel
344 quadrilateral results, and different vowels plot in well-separated regions of this acoustic space
345 (³⁰ p. 161). A vowel quadrilateral for the synthetic vowels produced via the plate-type model
346 is shown in Figure 7. This plot confirms that the acoustic properties of the synthetic sam-
347 ples are broadly consistent with the patterns of formants of natural vowels documented in
348 the prior literature, with all samples falling within the quadrilateral. Additionally, the sam-
349 ples locate to disparate regions of the quadrilateral, suggesting they may be perceived as
350 separable vowels.

351 Critical to the success of vowel production is whether or not the vowels are discriminable
352 and identifiable, that is whether or not they can be easily differentiated and transmit the
353 intended vowel to the listener, regardless of how non-overlapping their formant locations may
354 be in frequency space. These qualities were evaluated in an experimental program. First,
355 distances in perceptual space between the stimuli were obtained by asking participants to
356 rate inter-stimuli dissimilarity for all possible pairings. A multidimensional scaling analysis
357 was performed on the distances, which could be mapped to a two-dimensional projection
358 with minimal stress, in order to establish if the five synthetic sounds occupy discernibly
359 different regions in perceptual space. A vowel classification experiment was then carried out
360 to assess vowel identity and its consistency both within and between individuals.

361 Vowels and their pronunciations have evolved considerably since the time of Grosseteste,
362 and it goes without saying that we were unable to run experiments with participants with
363 a medieval language background. However, it is reasonable to expect that the mechanisms

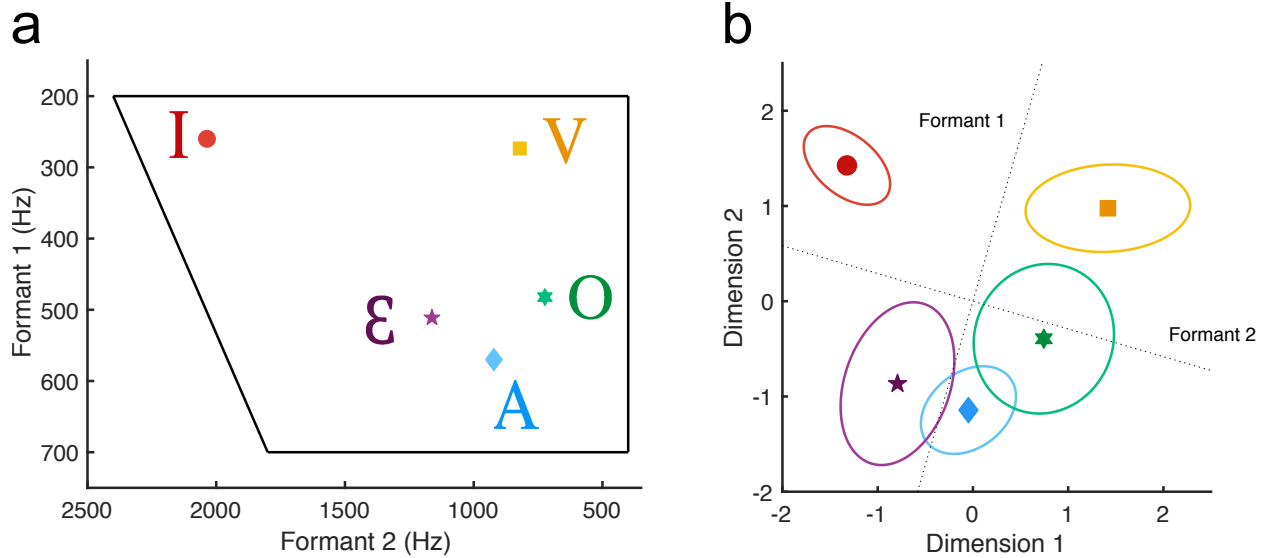


FIG. 7. (a) Acoustic map of the recorded synthetic vowels based on their measured first and second formant (F_1 and F_2) frequencies. The quadrilateral indicates the area within which discernible vowels are expected from previous literature³⁰. Blue diamond = sample A, purple pentagram = sample \mathcal{E} , red circle = sample I, green hexagram = sample O, orange square = sample V. (b) Scatter plot of MDS analysis for the perception of the same five synthetic vowels. Mappings were averaged across participants after Procrustes realignment. The mean locations for each sample are shown, with ellipses representing 1 SD of bivariate normal distributions fitted to the data. Interpretative axes were obtained by Procrustes analysis with the data from (a), and plotted as dotted lines.

364 of vowel perception have broadly remained constant to the modern era, although some finer
 365 elements of speech perception vary as a result of differing cultural and language contexts³¹.
 366 For this reason, we selected participants from a range of language backgrounds.

367 C. Multidimensional scaling experiment

368 In the first psychophysical experiment the five stimuli were presented to both native
369 and non-native English speakers to obtain dissimilarity scores. The *.wav* files (sampling
370 rate 44,100 Hz, 16 bit, monophonic) were all normalized to 0 dB relative to full scale and
371 limited to a duration of 1.70 s in Audacity, to be played through a pair of Sennheiser
372 HD201 Closed Dynamic Stereo headphones. The experiment was built using the open-
373 source Matlab function set Psychtoolbox³², and run using the same laptop and headphones
374 in quiet conditions. 20 participants took part in the experiment (12 female, 8 male, mean
375 age 25 years). Participants were asked for their country of origin (13 UK, 1 USA, 2 India,
376 2 Bulgaria, 1 Germany, 1 Poland), if they were native or non-native English speakers and if
377 non-native what their native language was (16 native English speakers [13 monolingual UK,
378 1 monolingual USA, 2 bilingual in English and Hindi], 4 non-native [2 Bulgarian, 1 German,
379 1 Polish]).

380 Participants were first played each of the five stimuli once for familiarity. Pairs of record-
381 ings were then presented separated by a 300 ms pause, and participants registered their
382 perceived dissimilarity via a keyboard, from 0 (identical) to 7 (very dissimilar). For stimuli
383 $i, j = 1, \dots, 5$, all possible pairs were presented once in a random order, for both (i, j) and
384 (j, i) sequences, to give a dissimilarity response matrix. From this, a symmetric matrix was
385 constructed for each participant by taking means of (i, j) and (j, i) values. For six of the
386 participants a single set of dissimilarity judgments was collected, while 14 went through the

387 experiment twice. Since no systematic differences in dissimilarity scores were found between
388 repeats, their symmetric matrices were averaged.

389 Kruskal’s non-metric multidimensional scaling (MDS)³³ was performed on the symmet-
390 rical matrices to approximate the relative locations in perceptual space of the samples for
391 individual participants. Once Euclidean coordinates were obtained from MDS analysis,
392 these were plotted to inspect their agreement with the formant plots of the samples. Visual
393 inspection of the mappings showed a clear correspondence between the first dimension of
394 scaling and F_2 , and the second dimension of scaling and F_1 , for the majority of participants,
395 which was later formally analyzed as described below. This agrees with previous studies
396 that find human vowel discrimination primarily tracks the frequency position of F_2 , which
397 corresponds to perceived vowel advancement, and secondarily tracks the frequency position
398 of F_1 , corresponding to perceived vowel height³⁴. There were four exceptions for this agree-
399 ment; notably, these data sets were from the four non-native English speaking participants.
400 Further inspection showed that these data agreed with F_2 and F_1 when plotted in the first
401 and third dimension from the MDS, respectively, and hence these mappings were taken
402 forward in the analysis.

403 Data sets then underwent Procrustes analysis, which permitted similarity transformations
404 of the mappings (uniform scaling, orthogonal rotation, translation and reflection) in order to
405 give the best concordance across participants while maintaining relative perceptual distances
406 within mappings³⁵. Once realigned, data sets were analyzed to extract the statistics for each
407 stimulus as located in perceptual space by participants. Figure 7.b shows the mean positions
408 for each stimulus, plotted as solid symbols. Ellipses show one standard deviation of the

409 bivariate distribution of each vowel within the two dimensions of scaling. Sample O gave
410 rise to the most spread compared to the other vowels, indicating that participants differed
411 most in where to locate it in their perceptual space, relative to the other vowels. This is
412 likely related to the strong degree of variation present in open back vowel pronunciations
413 across dialects of English.

414 Procrustes analysis was also performed between the realigned perceptual space data and
415 the acoustics-based vowel quadrilateral generated from formant data, in order to obtain
416 axes for interpretation of the MDS analysis, labelled as ‘Formant 1’ and ‘Formant 2’. The
417 distribution of relative perceptual locations for the five synthetic samples (Figure 7 b) show
418 a clear agreement with their placing in the F_2/F_1 frequency space (Figure 7 a), primarily
419 with the samples occupying separate (i.e. discriminable) regions in perceptual space, albeit
420 with some overlap between participants.

421 Monte Carlo simulations were carried out to evaluate the likelihood of stimuli being
422 mapped to distinct regions due to chance, and consistently with the same relative orientation.
423 From 26 simulations, only 20 generated data that could be mapped by MDS. After Procrustes
424 analysis of these 20 mappings, none gave rise to a distinct region for any of the stimuli (i.e.
425 non-overlapping regions bound by one standard deviation of stimuli mean position), and all
426 stimuli regions had an area above 5 scaling space units², compared to a mean of 1.2 scaling
427 space units² for participant-generated data. For all mappings, shown in Figure 9 in the
428 Supplementary Material, the relative orientation of vowels were different. A more extensive
429 simulation was carried out to generate 100 mappings, whose ellipses had a mean of 7 scaling
430 space units², shown in Figure 10. We therefore conclude that the results of mapping the

431 participant data, with stimuli occupying separable regions and a relative orientation in
432 agreement with the acoustic analysis, are not owing to chance.

433 **D. Vowel classification experiment**

434 Fourteen of the participants (ten native English speakers; four non-native English speak-
435 ers) also completed a second test, to obtain vowel classifications for the stimuli. Participants
436 were asked to listen to the recordings with headphones and assign them labels which best
437 agreed with their percepts. Participants were not expected to be familiar with IPA notation,
438 instead selecting one of the following options: “‘ah’ as in spa”, “‘eh’ as in get”, “‘ee’ as in
439 beat”, “‘o’ as in cot”, or “‘oo’ as in zoo”; corresponding to /ɑ, ɛ, i, ə, u/, respectively. These
440 options are also summarized in Table I in the Supplementary Material. Each stimulus ap-
441 peared in a familiarization phase once in this order, followed by a test phase in which they
442 were presented a further four times in a randomized order.

443 Responses from the familiarization phase were not included in the analysis, as participants
444 had not heard all of the vowels at that time. The data from individual participants did not
445 show any correlation between classification confusions and being a native/non-native English
446 speaker, which is not surprising given the coarseness of the classification system. Figure 8
447 shows the distributions of responses for each stimulus, with pie charts for each stimulus
448 being centered at the stimulus’ position in acoustic space as calculated above. The data are
449 also given in Table III in the Supplementary Material.

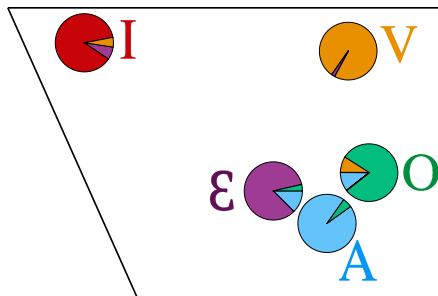


FIG. 8. Classifications obtained for each of the five samples from the second listening test. The pie charts for each sample, showing participants’ classifications, are centered at the samples’ locations when mapped in acoustic space, as shown in Figure 7 a. Responses are indicated by color: “‘ah’ as in spa” (/ɑ/) in blue, “‘eh’ as in get” (/ɛ/) in purple, “‘ee’ as in beat” (/i/) in red, “‘o’ as in cot” (/ɔ/) in green, and “‘oo’ as in zoo” (/u/) in orange.

450 **E. Results: MDS and classification experiments**

451 Listening to isolated vowels is not a common activity in daily life, and listening to isolated
 452 vowels without having any reference to the speaker is also unusual. In addition, these stimuli
 453 are clearly non-human in origin given the identical electrolarynx acoustic input in each case.
 454 Some confusion is therefore inevitable. As may be expected, the synthetic vowel with the
 455 broadest spread of placement in perceptual space (indicated by its ellipse in Figure 7 b having
 456 the greatest area) was also the least reliably classified sound, Sample O, which received 80.4%
 457 correct classifications and 10.7% and 8.9% misclassifications as ‘ah’ and as ‘oo’ respectively.
 458 The greatest source of misclassification was the assigning of both Sample \mathcal{E} and Sample O
 459 as ‘ah’ (12.5% and 10.7% respectively). The perceptual space generated by MDS analysis
 460 and the acoustic space from formant data both show Sample E and Sample O located in
 461 close proximity to Sample A, which itself was classified as ‘ah’ with high agreement. Indeed,

462 on the perceptual map these are the only two instances of overlapping standard deviations
463 from the samples' means. It can be said with confidence that the samples are perceived,
464 imperfectly, as vowels, spanning a large proportion of vowel perceptual space.

465 As well as the samples being consistently classified by participants, these classifications
466 were overwhelmingly in accordance with the mapping specified in the *DGS*, according to
467 which the vocal tract models were constructed, when these five vowel letters are related to
468 phonemes, as given in Table I in the Supplementary Material. Of course, we cannot be sure
469 that Grosseteste would have had these same phonetic sounds in mind (namely 'A' mapped
470 to /a/, 'E' mapped to /ε/, 'I' mapped to /i/, 'O' mapped to /ɔ/, and 'V' mapped to /u/).
471 The classification task did not test for exact identity between stimuli and labels; participants
472 were asked to select the closest match from the five options given rather than provide their
473 own labels. However, it is worth stating that as there are 120 possible permutations of
474 mapping five labels to five stimuli ($P(5) = 5! = 120$), it would be unlikely to observe this
475 specific mapping by chance alone across numerous participants. We can therefore conclude
476 that the shapes Grosseteste specified for shaping the vocal tract during vowel production
477 are compatible with their related phonemes when present in the mouth end of the vocal
478 tract (or other acoustic chamber), in a five-vowel system.

479 IV. DISCUSSION

480 While sometimes described as a scientist, and undoubtedly instrumental in the conception
481 of the scientific experimental method³⁶, we must be careful when reading Grosseteste's treatises
482 not to impute any sense of experimental or even observational basis for his theories,

483 however elegant the logical or mathematical arguments found therein. Recent interdisci-
484 plinary research has found that the origin of such theories, though they may be wrong
485 within the context of current scientific understanding, may still best be explained as result-
486 ing from direct observation, such as for his novel theory of rainbow formation⁸. However
487 others, though they may have been correct, are unlikely to have had a direct observational
488 basis, such as his three-dimensional theory of colour space as expressed in the *De colore*⁶.
489 These works remain remarkable achievements, and the desire to mathematicize the mental
490 or material world was a fundamental evolution for intellectual history in the medieval and
491 early modern era.

492 In his treatise on sound, Grosseteste is applying a similar mathematical framework of
493 combinatorics as his theory of colour, but to vowels. There are, however, some interesting
494 differences between the two. In the *De colore*, Grosseteste is clear that colour space is
495 continuous, as he describes the infinite ‘diminutions’ between the extrema of the space. That
496 he constructs the parameter space to reflect established intuitions about space and distance
497 is therefore quite sensible; colours are connected along routes, which may be traversed by
498 increasing or decreasing one, two, or all three of the space’s parameters. This particular
499 feature of the theory we can presume was likely based on direct observation, and the subtle
500 and continuous variations in colours seen in the world and, explicitly, in rainbows. In the
501 *De generatione sonorum*, Grosseteste again constructs a generative scheme to account for
502 the variety within a perceptual phenomenon, but it is this time categorical and discrete,
503 accounting for the varieties of vowels and their external representational forms, letters.

504 The scheme is defined by what he says are the three types of simple, self-similar move-
505 ments: linear, circular, and dilational-constrictional. These simple movements may be com-
506 bined, but only a subset yield novel categories of movement: combining linear with circu-
507 lar, linear with dilational-constrictional, circular with circular, and circular with dilational-
508 constrictional. These descriptions of movement are readily interpreted as the three types of
509 geometric similarity transformations—translation, rotation, and uniform scaling (with re-
510 flection being equivalent to rotation through a higher dimension)—though it should be noted
511 that no diagrams are found in extant manuscripts, and this is just one possible interpretive
512 scheme³⁷. The treatise can be read as one primarily about types of movement, and relies
513 heavily on the false premise that sounds of different qualities are discriminable based on the
514 category of vibrational movement, rather than the spectral filtering achieved by differently-
515 shaped acoustic chambers with varying resonant frequencies, and other language-specific
516 factors. Although this theory is mistaken about the underlying source of vowel timbre,
517 Grosseteste nevertheless constructs an elegant theory that attempts to account for the cat-
518 egorical nature of vowel perception, and the representation of vowels as letters.

519 Reading this text today prompts us to examine what may constitute permissible evidence
520 in science. For Grosseteste, the shapes of letters could serve as the primary evidence for
521 his claims regarding the shape of the vocal tract, and the forms of mental representations
522 of vowels; within the medieval paradigms of Aristotelian mimesis and the liberal arts, this
523 was a scientifically orthodox and justifiable use of observations to infer properties of the
524 natural world. Although we do not share these paradigms as modern scientists, we share
525 in the methodological framework of setting our own standards for permissible evidence; in

526 many cases such sources of evidence are far-removed from the phenomenon we attempt
527 to study. A generous reading of the *DGS* could be that Grosseteste is engaged in mod-
528 elling; do abstract movement categories offer a viable framework for the robust, categorical
529 representation and perception of speech sounds, despite their continuous variety and noisy
530 instances? Although our models of speech processing have matured in their awareness of
531 acoustics and physiology^{38–40}, they share the underlying goal of understanding how speech
532 signals are processed and represented.

533 The *DGS* does make strong claims about the morphology of the vocal tract during vowel
534 production, which are clearly incorrect in asserting the presence of geometric shapes. How-
535 ever, we have shown, through artificial vowel synthesis and the methods of spectral analysis
536 and psychophysical testing of vowel perception, that these geometric shapes can in fact be
537 incorporated at the mouth end of acoustic chambers that give rise to discriminable vowel
538 sounds. This is plausibly due to degree of freedom present in the remainder of the acoustic
539 chamber, i.e. the laryngeal and pharyngeal cavity, and the many-to-one property of acous-
540 tic chambers and their spectral output⁴¹, meaning that unique speech sounds may have
541 multimodal or highly nonlinear mappings in articulator space⁴². In the thirteenth-century
542 Grosseteste would only have had visual and proprioceptive measurements of the lips, teeth
543 and tongue, so any requirements of the rest of the vocal tract for vowel production could
544 not have impacted his theory.

545 How influential the *DGS* was on the developing field of phonetics is difficult to say. Roger
546 Bacon, a student of Grosseteste’s who praised his mathematical approach to understanding
547 nature, describes similar notions of relating the number of vowels in languages to the number

548 of fundamental classes of movements in his text on Greek Grammar⁴³. However, he seems
549 to criticize these theories as falling outside the scope of the ‘pure grammarian’, instead they
550 should be left to the disciplines of metaphysics and of music⁴⁴. Specifically, he is engaging
551 with the content of the *Tractatus de grammatica*. Circulating at the time, the anonymous
552 *Tractatus* was widely attributed to Aristotle, but Bacon shows this to be unjustified, and
553 the treatise was later sometimes ascribed to Grosseteste.

554 Readers familiar with Hangul, the native Korean alphabet devised by King Sejong the
555 Great (1397-1450) in the fifteenth century, may find similarities between Grosseteste’s theory
556 of non-arbitrary letter shapes and the apparent similarity between Hangul consonant forms
557 and their corresponding places of articulation⁴⁵. However, we have no record of a reception
558 of Grosseteste’s work in east Asia, and any direct connection seems improbable. Moreover,
559 while the articulatory basis of the Hangul alphabet is often stated as matter of fact, and has
560 been written about since only a few years after Hangul was devised (such as in *Hwunmin*
561 *cengum haylyey* [Explanations and Examples of the Correct Sounds for the Instruction of the
562 People], published in 1446), there are competing theories. It seems equally likely that Hangul
563 consonants were instead influenced by or modelled on the Mongol ’Phags-pa alphabet, itself
564 derived from Tibetan, as suggested by Keith Whinnom⁴⁶. It could therefore be the case that
565 in Hangul and its reception we find a thesis parallel to claims made in the *DGS*: the notion
566 of glyph iconicity being used as a kind of pedagogical or philosophical device to explain their
567 forms.

568 Theories attempting to draw direct relationships between the shaping of articulators and
569 the shapes of letters surfaced again in the seventeenth century, with Franciscus Mercurius

570 van Helmont claiming that intrinsic to the Hebrew alphabet was found a phonetic guide
571 to its pronunciation⁴⁷, and Bishop John Wilkins attempting to construct a visual alphabet
572 of speech sound diagrams⁴⁸. In neither case is there an explicit connection to the *DGS*.
573 Such theories relating letter shapes to vocal tract shapes paved the way for the speaking
574 machine of Wolfgang von Kempelen in 1780, and, later, the set of ‘visible speech’ symbols
575 by Alexander Melville Bell^{49,50}.

576 Lastly, an essay published in 1772 by Charles Davy makes near identical claims regarding
577 the representations of the vocal tract in the letter shapes of vowels⁵¹ (p84-87), but again,
578 any connection to Grosseteste’s theory is not made explicit and may be entirely accidental.
579 It should also be noted that Davy’s text was not written as a serious scientific endeavour but
580 as an amusing romp through classical trivia, with Davy himself writing: “The Editor will
581 not undertake to defend it: as a *whimsical* conjecture, it may still afford some entertainment.
582 *Better* reasons might perhaps be offered in its favour than what appear at present”, before
583 stating his belief that the Greeks’ visual representation of the vocal tract in letter shapes is
584 what enabled their literary success. It may simply be the case that such theories were best
585 appreciated as a form of intellectual entertainment, rather than serious scientific endeavour.
586 Now, with the advent of recent studies into glyph iconcity^{17,18}, theories of non-arbitrary
587 representation of letter shapes are again being considered, albeit from a more nuanced and
588 experimental standpoint.

589 V. CONCLUSION

590 In the treatise *De generatione sonorum* [*On the Generation of Sounds*], Robert Gros-
591 seteste attempts a mathematicization of the perceptual space of vowels. With this paper
592 we show that the treatise formulates vowels—their production, perception, and representa-
593 tions both mental and in writing—into a coherent framework of geometric figures, which
594 are combinatorially generated from basic types of movement. Although clearly incorrect in
595 his understanding of vocal acoustics, and ignorant of the supporting physiology, Grosseteste
596 shows remarkable insight in his approach to explaining why vowels are categorical in nature,
597 and how auditory, visual, and motor faculties play complementary roles in speech percep-
598 tion. His theory touches on principles highly relevant to contemporary neuroscience, namely
599 the nature of mental representations and their relationship to external stimuli, and the inte-
600 gration of different sensory faculties. Finally, aspects of Grosseteste’s theory of speech can
601 be expressed in a scientific, falsifiable manner, which we show here to have been potentially
602 commensurable with the sensory data available at the time.

603 ACKNOWLEDGMENTS

604 This work was supported by the AHRC under grant AH/N001222/1. JSH holds a DPhil
605 studentship from the Andrew W. Mellon Foundation, hosted by the Oxford interdisciplinary
606 research center TORCH. HES was supported by a Visiting Fellowship from the Institute
607 of Advanced Study at Durham University. The work presented here emerged from the

608 collaborative Ordered Universe Project (<http://ordered-universe.com/>), which focuses on
609 interdisciplinary readings of the scientific works of Robert Grosseteste (c.1170–1253).

610 The authors would like to thank all the participants of the collaborative workshops during
611 which the *DGS* was discussed [2nd-3rd October 2014, “13th Century Science in a Multi-
612 Disciplinary Perspective”, Pembroke College, Oxford (funded by the Mahfouz Foundation);
613 8th-10th April 2015 “Knowing and Speaking: On the Generation of Sounds and On the
614 Liberal Arts”, Bishop Grosseteste University, Lincoln; 25th-28th November 2015, “On the
615 Liberal Arts and On the Generation of Sounds: Robert Grosseteste’s Early Treatises and
616 Their Reception”, Durham University]. We thank John Coleman for his advice on RP,
617 Middle English, and Latin pronunciation, the use of IPA symbols, and the origins of the
618 Hangul alphabet. We appreciate Brian Tanner’s contributions to the discussion of sound
619 and the movements of the vocal tract, Neil Lewis’s suggestions on issues of translation
620 and interpretation of the medieval texts, and Cecilia Panti’s discussions on the text. We
621 also thank the two anonymous reviewers for their comments on an earlier edition of this
622 manuscript. Finally, the authors thank all the participants who took part in the listening
623 tests.

624 **APPENDIX: SUPPLEMENTARY MATERIAL**

Letter shape	Phoneme	Example
A	/ɑ/	‘ah’ as in ‘part’
ℰ	/ɛ/	‘eh’ as in ‘get’
I	/i/	‘ee’ as in ‘beat’
O	/ɔ/	‘o’ as in ‘cot’
V	/u/	‘oo’ as in ‘zoo’

TABLE I. Our interpretation of phonemes from the vowel letters Grosseteste uses in *DGS*. The third column also shows the options given to participants in the classification listening test.

	larynx lips															
Sample A	22	8	18	8	8	12	16	20	24	26	28	30	32	34	38	24
Sample E	12	8	12	8	22	14	14	10	16	24	18	10	16	24	18	10
Sample I	16	32	32	32	32	30	30	20	12	12	8	8	8	10	10	10
Sample O	8	20	12	12	12	10	8	8	16	24	30	32	30	24	16	10
Sample V	8	32	10	8	30	28	26	24	22	20	18	16	14	12	10	8

TABLE II. Diameters (in mm) of the employed plate-type model of Arai *et al.*²⁹ used to create the tracts shown in Figure 5 and to synthesize the five speech sounds (Sample A, Sample E, Sample I, Sample O, Sample V) based on Grosseteste’s five movement types.

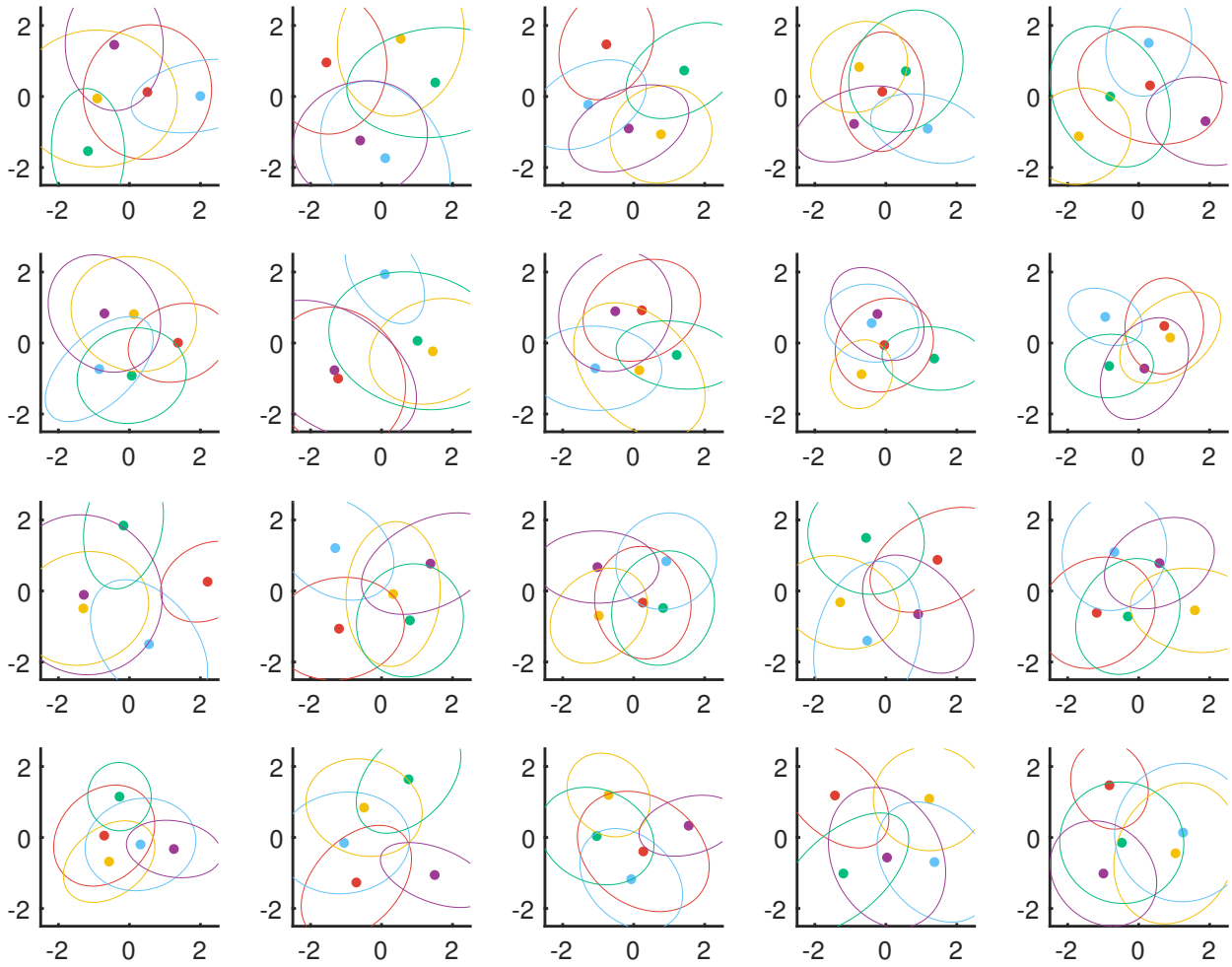


FIG. 9. 20 examples of Monte Carlo simulations that generated data sets for which a MDS mapping was possible. No simulation produced dissimilarity data that when mapped featured a distinct area for a stimulus, as bound by one standard deviation from its mean position (indicated by ellipses).

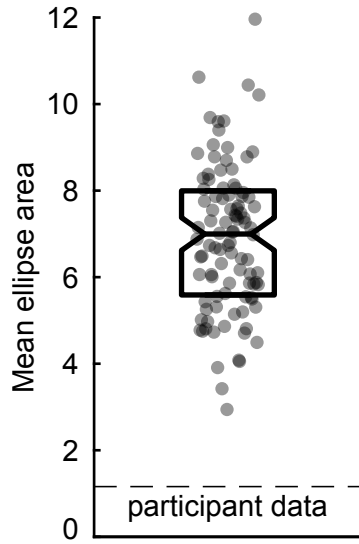


FIG. 10. The results of 100 Monte Carlo simulations of the MDS experiment. The mean ellipse areas from each simulation (which comprised 20 randomized participant data sets) are shown. The box plot indicates the mean and quartiles of the distribution, with a 95% confidence interval on the mean shown as a notch. The mean of the participant data set is indicated by a dashed line.

‘ah’ as in ‘part’ ‘eh’ as in ‘get’ ‘ee’ as in ‘beat’ ‘o’ as in ‘cot’ ‘oo’ as in ‘zoo’

Sample A	64	0	0	6	0
Sample E	8	59	0	3	0
Sample I	0	8	59	0	3
Sample O	7	1	0	57	5
Sample V	0	1	0	1	68

TABLE III. Results from the classification experiment ($N = 14$). Each participant classified each sample five times, choosing from the five possible responses in the top row of the table.

625 **References**

626 ¹C. Burnett, D. C. Lindberg, and M. H. Shank, “Translation and Transmission of Greek
627 and Islamic Science to Latin Christendom,” in *The Cambridge History of Science*, Vol. 2
628 of *The Cambridge History of Science* (Cambridge University Press, 2013), pp. 341–364,
629 doi: [10.1017/CH09780511974007.016](https://doi.org/10.1017/CH09780511974007.016).

630 ²C. Burnett, “The Introduction of Aristotle’s Natural Philosophy into Great Britain : A
631 Preliminary Survey of the Manuscript Evidence,” in *Aristotle in Britain during the Middle*
632 *Ages*, Vol. 5 of *Rencontres de Philosophie Mdivale* (Brepols Publishers, 1996), pp. 21–
633 50, <https://www.brepolonline.net/doi/abs/10.1484/M.RPM-EB.4.000076>, doi: [10.](https://doi.org/10.1484/M.RPM-EB.4.000076)
634 [1484/M.RPM-EB.4.000076](https://doi.org/10.1484/M.RPM-EB.4.000076).

635 ³P. De Leemans, “Aristotle Transmitted: Reflections on the Transmission of Aristotelian
636 Scientific Thought in the Middle Ages,” *International Journal of the Classical Tradition*
637 **17**(3), 325–353 (2010) <https://doi.org/10.1007/s12138-010-0200-9> doi: [10.1007/](https://doi.org/10.1007/s12138-010-0200-9)
638 [s12138-010-0200-9](https://doi.org/10.1007/s12138-010-0200-9).

639 ⁴This new critical edition and translation incorporates three manuscripts not known by the
640 last editor, Baur⁵².

641 ⁵G. E. M. Gasper, C. Panti, T. C. B. McLeish, and H. E. Smithson, eds., *Knowing and*
642 *Speaking: Robert Grossetestes De artibus liberalibus On the Liberal Arts and De genera-*
643 *tione sonorum On the Generation of Sounds*, Vol. 1 of 6 of *The Scientific Works of Robert*
644 *Grosseteste* (Oxford University Press, Oxford, 2019), the critical edition and English trans-
645 lation of *On the Generation of Sounds* which form chapter 11 of this volume was provided

646 by S. O. Sønnesyn.

647 ⁶H. E. Smithson, G. Dinkova-Bruun, G. E. M. Gasper, M. Huxtable, T. C. B. McLeish,
648 and C. Panti, “A three-dimensional color space from the 13th century,” *Journal of the*
649 *Optical Society of America. A, Optics, image science, and vision* **29**(2), A346–A352 (2012)
650 <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3287286/>.

651 ⁷H. E. Smithson, P. S. Anderson, G. Dinkova-Bruun, R. A. E. Fosbury, G. E. M. Gasper,
652 P. Laven, T. C. B. McLeish, C. Panti, and B. Tanner, “A color coordinate system from a
653 13th century account of rainbows,” *Journal of the Optical Society of America. A, Optics,*
654 *image science, and vision* **31**(4), A341–A349 (2014) [http://www.ncbi.nlm.nih.gov/pmc/](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4381719/)
655 [articles/PMC4381719/](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4381719/).

656 ⁸J. S. Harvey, H. E. Smithson, C. R. Siviour, G. E. M. Gasper, S. O. Sønnesyn, B. K. Tanner,
657 and T. C. B. McLeish, “Bow-shaped caustics from conical prisms: a 13th-century account
658 of rainbow formation from Robert Grossetestes *De iride*,” *Applied Optics* **56**(19), G197
659 (2017) <https://www.osapublishing.org/abstract.cfm?URI=ao-56-19-G197>, doi: 10.
660 [1364/AO.56.00G197](https://www.osapublishing.org/abstract.cfm?URI=ao-56-19-G197).

661 ⁹R. G. Bower, T. C. B. McLeish, B. K. Tanner, H. E. Smithson, C. Panti, N. Lewis,
662 and G. E. M. Gasper, “A Medieval Multiverse: Mathematical Modelling of the 13th
663 Century Universe of Robert Grosseteste,” *Proceedings of the Royal Society A: Mathe-*
664 *matical, Physical and Engineering Sciences* **470**(2167), 20140025–20140025 (2014) [http:](http://arxiv.org/abs/1403.0769)
665 [//arxiv.org/abs/1403.0769](http://arxiv.org/abs/1403.0769), doi: 10.1098/rspa.2014.0025.

666 ¹⁰J. C. Le Blon, *Coloritto, or, The harmony of colouring in painting* (1720), [https://](https://library.si.edu/digital-library/book/colorittoharmon00lebl)
667 library.si.edu/digital-library/book/colorittoharmon00lebl.

668 ¹¹Young Thomas, “II. The Bakerian Lecture. On the theory of light and colours,”
669 *Philosophical Transactions of the Royal Society of London* **92**, 12–48 (1802) [https://](https://royalsocietypublishing.org/doi/abs/10.1098/rstl.1802.0004)
670 royalsocietypublishing.org/doi/abs/10.1098/rstl.1802.0004, doi: [10.1098/](https://doi.org/10.1098/rstl.1802.0004)
671 [rstl.1802.0004](https://doi.org/10.1098/rstl.1802.0004).

672 ¹²H. Chang, “Who cares about the history of science?,” *Notes and Records of the Royal So-*
673 *ciety of London* **71**(1), 91–107 (2017) [https://www.ncbi.nlm.nih.gov/pmc/articles/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5311899/)
674 [PMC5311899/](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5311899/), doi: [10.1098/rsnr.2016.0042](https://doi.org/10.1098/rsnr.2016.0042).

675 ¹³Although Grosseteste only refers explicitly to the shapes of letters, when read in the
676 broader context of the *DGS* and its focus on motion, a valid interpretation is that the
677 movements of handwriting played a central role in his thinking.

678 ¹⁴A comprehensive discussion of the seven liberal arts, and Grosseteste’s treatise on them,
679 may be found in⁵.

680 ¹⁵Aristotle, *Complete Works of Aristotle, Volume 1: The Revised Oxford Translation: Re-*
681 *vised Oxford Translation v. 1* (Princeton University Press, Princeton, NJ, 1984), ed.
682 Barnes, J.

683 ¹⁶A deeper dive into the intellectual influences for Grosseteste’s *DGS* can be found in⁵³.

684 ¹⁷D. S. Schmidtke, M. Conrad, and A. M. Jacobs, “Phonological iconicity,” *Frontiers*
685 *in Psychology* **5** (2014) [http://journal.frontiersin.org/article/10.3389/fpsyg.](http://journal.frontiersin.org/article/10.3389/fpsyg.2014.00080/abstract)
686 [2014.00080/abstract](http://journal.frontiersin.org/article/10.3389/fpsyg.2014.00080/abstract), doi: [10.3389/fpsyg.2014.00080](https://doi.org/10.3389/fpsyg.2014.00080).

- 687 ¹⁸N. Turoman and S. J. Styles, “Glyph guessing for oo and ee: spatial frequency information
688 in sound symbolic matching for ancient and unfamiliar scripts,” Royal Society Open Sci-
689 ence **4**(9), 170882 (2017) [http://rsos.royalsocietypublishing.org/lookup/doi/10.](http://rsos.royalsocietypublishing.org/lookup/doi/10.1098/rsos.170882)
690 [1098/rsos.170882](http://rsos.royalsocietypublishing.org/lookup/doi/10.1098/rsos.170882), doi: [10.1098/rsos.170882](https://doi.org/10.1098/rsos.170882).
- 691 ¹⁹B. Galantucci, C. A. Fowler, and M. T. Turvey, “The motor theory of speech perception
692 reviewed,” Psychonomic bulletin & review **13**(3), 361–377 (2006) [http://www.ncbi.nlm.](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2746041/)
693 [nih.gov/pmc/articles/PMC2746041/](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2746041/).
- 694 ²⁰K. Watkins, A. Strafella, and T. Paus, “Seeing and hearing speech excites the mo-
695 tor system involved in speech production,” Neuropsychologia **41**(8), 989–994 (2003)
696 <http://linkinghub.elsevier.com/retrieve/pii/S0028393202003160>, doi: [10.1016/](https://doi.org/10.1016/S0028-3932(02)00316-0)
697 [S0028-3932\(02\)00316-0](https://doi.org/10.1016/S0028-3932(02)00316-0).
- 698 ²¹F. Pulvermuller, M. Huss, F. Kherif, F. Moscoso del Prado Martin, O. Hauk, and Y. Shty-
699 rov, “Motor cortex maps articulatory features of speech sounds,” Proceedings of the Na-
700 tional Academy of Sciences **103**(20), 7865–7870 (2006), doi: [10.1073/pnas.0509989103](https://doi.org/10.1073/pnas.0509989103).
- 701 ²²I. G. Meister, S. M. Wilson, C. Deblieck, A. D. Wu, and M. Iacoboni, “The Essential Role
702 of Premotor Cortex in Speech Perception,” Current Biology **17**(19), 1692–1696 (2007)
703 <http://linkinghub.elsevier.com/retrieve/pii/S0960982207019690>, doi: [10.1016/](https://doi.org/10.1016/j.cub.2007.08.064)
704 [j.cub.2007.08.064](https://doi.org/10.1016/j.cub.2007.08.064).
- 705 ²³M. Sato, P. Tremblay, and V. L. Gracco, “A mediating role of the premotor cortex in
706 phoneme segmentation,” Brain and Language **111**(1), 1–7 (2009) [http://linkinghub.](http://linkinghub.elsevier.com/retrieve/pii/S0093934X09000418)
707 [elsevier.com/retrieve/pii/S0093934X09000418](http://linkinghub.elsevier.com/retrieve/pii/S0093934X09000418), doi: [10.1016/j.bandl.2009.03](https://doi.org/10.1016/j.bandl.2009.03).

708 002.

709 ²⁴A. DAusilio, I. Bufalari, P. Salmas, and L. Fadiga, “The role of the motor sys-
710 tem in discriminating normal and degraded speech sounds,” *Cortex* **48**(7), 882–887
711 (2012) <https://linkinghub.elsevier.com/retrieve/pii/S0010945211001614>, doi:
712 [10.1016/j.cortex.2011.05.017](https://doi.org/10.1016/j.cortex.2011.05.017).

713 ²⁵R. Mottonen and K. E. Watkins, “Motor Representations of Articulators Contribute to
714 Categorical Perception of Speech Sounds,” *Journal of Neuroscience* **29**(31), 9819–9825
715 (2009), doi: [10.1523/JNEUROSCI.6018-08.2009](https://doi.org/10.1523/JNEUROSCI.6018-08.2009).

716 ²⁶X. Tian and D. Poeppel, “Mental imagery of speech: linking motor and perceptual systems
717 through internal simulation and estimation,” *Frontiers in Human Neuroscience* **6** (2012)
718 <https://www.frontiersin.org/articles/10.3389/fnhum.2012.00314/full> doi: [10.](https://doi.org/10.3389/fnhum.2012.00314)
719 [3389/fnhum.2012.00314](https://doi.org/10.3389/fnhum.2012.00314).

720 ²⁷M. H. Bakalla, *Ibn Jinni, an early Arab Muslim phonetician: an interpretive study of*
721 *his life and contribution to linguistics*, limited ed. ed. (European Language Publications,
722 Taipei, 1982).

723 ²⁸R. Allott, “The articulatory basis of the alphabet,” in *Becoming Loquens: More Studies*
724 *in Language Origins*, edited by B. H. Bichakjian, T. Chernigovskaya, A. Kendon, and
725 A. Moller, Vol. 1 of *Bochum Publications in Evolutionary Cultural Semiotics*, 1 ed. (Peter
726 Lang, Frankfurt am Main, 2000), p. 18.

727 ²⁹T. Arai, N. Usuki, and Y. Murahara, “Prototype of a vocal-tract model for vowel produc-
728 tion designed for education in speech science.,” in *INTERSPEECH* (2001), pp. 2791–2794.

729 ³⁰J. C. Catford, *A practical introduction to phonetics*, Oxford textbooks in linguistics, 2nd
730 ed ed. (Oxford University Press, Oxford ; New York, 2001).

731 ³¹P. K. Kuhl, S. Kiritani, T. Deguchi, A. Hayashi, E. B. Stevens, C. D. Dugger, and
732 P. Iverson, “Effects of language experience on speech perception: American and Japanese
733 infants perception of /ra/ and /la/,” *The Journal of the Acoustical Society of Amer-*
734 *ica* **102**(5), 3135–3136 (1997) <http://asa.scitation.org/doi/abs/10.1121/1.420646>,
735 doi: [10.1121/1.420646](https://doi.org/10.1121/1.420646).

736 ³²D. H. Brainard, “The Psychophysics Toolbox,” *Spatial Vision* **10**(4), 433–436 (1997)
737 https://brill.com/view/journals/sv/10/4/article-p433_15.xml, doi: [10.1163/](https://doi.org/10.1163/156856897X00357)
738 [156856897X00357](https://doi.org/10.1163/156856897X00357).

739 ³³J. B. Kruskal, “Nonmetric multidimensional scaling: A numerical method,” *Psychometrika*
740 **29**(2), 115–129 (1964) <https://link.springer.com/article/10.1007/BF02289694>,
741 doi: [10.1007/BF02289694](https://doi.org/10.1007/BF02289694).

742 ³⁴J. M. Sinnott, C. H. Brown, W. T. Malik, and R. A. Kressley, “A multidimensional scaling
743 analysis of vowel discrimination in humans and monkeys,” *Perception & Psychophysics*
744 **59**(8), 1214–1224 (1997).

745 ³⁵R. Sibson, “Studies in the Robustness of Multidimensional Scaling: Procrustes Statistics,”
746 *Journal of the Royal Statistical Society. Series B (Methodological)* **40**(2), 234–238 (1978)
747 <http://www.jstor.org/stable/2984761>.

748 ³⁶A. C. Crombie, *Robert Grosseteste and the Origins of Experimental Science, 1100-1700*
749 (Clarendon Press, 1953).

750 ³⁷Other schemes of interpretation may be found in a forthcoming volume, which contains
751 the critical edition, translation, and interdisciplinary analyses of the *DGS*⁵⁴.

752 ³⁸J. L. McClelland and J. L. Elman, “The TRACE model of speech perception,” *Cogni-*
753 *tive Psychology* **18**(1), 1–86 (1986) [http://linkinghub.elsevier.com/retrieve/pii/](http://linkinghub.elsevier.com/retrieve/pii/0010028586900150)
754 [0010028586900150](http://linkinghub.elsevier.com/retrieve/pii/0010028586900150), doi: [10.1016/0010-0285\(86\)90015-0](https://doi.org/10.1016/0010-0285(86)90015-0).

755 ³⁹K. N. Stevens, “Toward a model for lexical access based on acoustic landmarks and dis-

756 tinctive features,” *The Journal of the Acoustical Society of America* **111**(4), 1872–1891
757 (2002) <http://asa.scitation.org/doi/10.1121/1.1458026>, doi: [10.1121/1.1458026](https://doi.org/10.1121/1.1458026).

758 ⁴⁰G. Hickok and D. Poeppel, “The cortical organization of speech processing,” *Nature Re-*
759 *views Neuroscience* **8**(5), 393–402 (2007) <http://www.nature.com/articles/nrn2113>,
760 doi: [10.1038/nrn2113](https://doi.org/10.1038/nrn2113).

761 ⁴¹B. S. Atal, J. J. Chang, M. V. Mathews, and J. W. Tukey, “Inversion of articulatorytoa-

762 coustic transformation in the vocal tract by a computersorting technique,” *The Journal*
763 *of the Acoustical Society of America* **63**(5), 1535–1555 (1978) [https://asa.scitation.](https://asa.scitation.org/doi/10.1121/1.381848)
764 [org/doi/10.1121/1.381848](https://asa.scitation.org/doi/10.1121/1.381848) doi: [10.1121/1.381848](https://doi.org/10.1121/1.381848).

765 ⁴²C. Qin and M. . Carreira-Perpin, “The geometry of the articulatory region that produces a

766 speech sound,” in *2009 Conference Record of the Forty-Third Asilomar Conference on Sig-*
767 *nals, Systems and Computers* (2009), pp. 1742–1746, doi: [10.1109/ACSSC.2009.5469741](https://doi.org/10.1109/ACSSC.2009.5469741).

768 ⁴³E. Nolan and S. A. Hirsch, *The Greek grammar of Roger Bacon and a fragment of his*
769 *Hebrew grammar*; (Cambridge University Press, 1902), [http://archive.org/details/](http://archive.org/details/cu31924021600790)
770 [cu31924021600790](http://archive.org/details/cu31924021600790).

- 771 ⁴⁴I. Roser-Catach, “Roger Bacon and Grammar,” in *Roger Bacon and the Sci-*
772 *ences: Commemorative Essays*, edited by J. Hackett (BRILL, 1997), google-Books-ID:
773 Gy3Vp7TurVUC.
- 774 ⁴⁵Y.-K. Kim-Renaud, *The Korean Alphabet: Its History and Structure* (University of Hawaii
775 Press, 1997), google-Books-ID: nonRl2cerIgC.
- 776 ⁴⁶E. R. Hope, “Letter Shapes in Korean Önmun and Mongol hPhagspa Alphabets,”
777 *Oriens* **10**(1), 150–159 (1957) <https://www.jstor.org/stable/1578766>, doi: 10.2307/
778 1578766.
- 779 ⁴⁷F. M. van Helmont, *Alphabeti veri naturalis hebraici brevissima delineatio* (1667).
- 780 ⁴⁸John Wilkins, *An Essay towards a Real Character and a Philosophical Lan-*
781 *guage* (Samuel Gellibrand, London, 1668), [http://archive.org/details/](http://archive.org/details/AnEssayTowardsARealCharacterAndAPhilosophicalLanguage)
782 [AnEssayTowardsARealCharacterAndAPhilosophicalLanguage](http://archive.org/details/AnEssayTowardsARealCharacterAndAPhilosophicalLanguage).
- 783 ⁴⁹A. M. Bell, *Visible speech: the science of universal alphabetic; or Self-interpreting physio-*
784 *logical letters, for the writing of all languages in one alphabet* (London, Simpkin, Marshall
785 & Co., 1867), <http://archive.org/details/visiblespeechsci00bellrich>.
- 786 ⁵⁰H. Dudley and T. H. Tarnoczy, “The Speaking Machine of Wolfgang von Kempelen,”
787 *The Journal of the Acoustical Society of America* **22**(2), 151–166 (1950) [https://asa.](https://asa.scitation.org/doi/abs/10.1121/1.1906583)
788 [scitation.org/doi/abs/10.1121/1.1906583](https://asa.scitation.org/doi/abs/10.1121/1.1906583) doi: 10.1121/1.1906583.
- 789 ⁵¹C. Davy, *Conjectural observations on the origin and progress of alphabetic writing*. (T.
790 Wright, London, 1772), <https://catalog.hathitrust.org/Record/001181056>.

791 ⁵²L. Baur, *Die philosophischen Werke des Robert Grosseteste, Bischofs von Lincoln* (Münster
792 i. W. : Aschendorff, 1912), <http://archive.org/details/diephilosophiede00baur>.

793 ⁵³S. O. Sønnesyn and G. E. M. Gasper, “Aristotle, Priscian, and Isidore,” in *Knowing*
794 *and Speaking: Robert Grosseteste’s De artibus liberalibus ‘On the Liberal Arts’ and De*
795 *generatione sonorum ‘On the Generation of Sounds’*, edited by G. E. M. Gasper, C. Panti,
796 T. C. B. McLeish, and H. E. Smithson, Vol. 1 of *The Scientific Works of Robert Grosseteste*
797 (Oxford University Press, Oxford, 2019).

798 ⁵⁴J. S. Harvey, R. C. White, H. E. Smithson, T. C. B. McLeish, D. Howard, and J. Coleman,
799 “Instrumental motions: Shaping and perceiving speech sounds,” in *Knowing and Speak-*
800 *ing: Robert Grosseteste’s De artibus liberalibus ‘On the Liberal Arts’ and De generatione*
801 *sonorum ‘On the Generation of Sounds’*, edited by G. E. M. Gasper, C. Panti, T. C. B.
802 McLeish, and H. E. Smithson, Vol. 1 of *The Scientific Works of Robert Grosseteste* (Oxford
803 University Press, Oxford, 2019), pp. 336–366.