Article:

Harvey, J. S., Smithson, H. E., Siviour, C. R. et al. (4 more authors) (2019) A thirteenth-century theory of speech. Journal of the Acoustical Society of America. pp. 937-947. ISSN 1520-8524

https://doi.org/10.1121/1.5119126

A thirteenth-century theory of speech

J. S. Harvey,[1,2] H. E. Smithson,[1] C. R. Siviour,[2] G. E. M. Gasper,[3,4] S. O. Sønnesyn,[3,4]
T. C. B. McLeish,[5,6] and D. M. Howard[7]

1) *Department of Experimental Psychology, University of Oxford*[a]

[2] *Department of Engineering Science, University of Oxford*

[3] *Department of History, Durham University*

[4] *Institute for Medieval and Early Modern Studies, Durham University*

[5] *Department of Physics, University of York*

[6] *Centre for Medieval Studies, University of York*

[7] *Department of Electronic Engineering, Royal Holloway,*
University of London

(Dated: 12 June 2019)

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

---

a) joshua.harvey@pmb.ox.ac.uk

I. INTRODUCTION

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

Robert Grosseteste (c.1175–1253) was a celebrated medieval thinker, who, as well as writing on philosophy and theology, developed an impressive corpus of treatises on the natural world. Here, we analyze one of these treatises—his text on sound and phonetics: *De generatione sonorum* [*On the Generation of Sounds*] (*DGS*). The *DGS* was probably written in the first decade of the thirteenth century, several centuries before the apparent 'scientific revolution in Early Modern Europe. It was a formative period, however, for the development of European scientific thought, during which the reception of Greek natural philosophy, enabled by their transmission, translations, and commentary from Arabic and Greek into Latin, prompted new conceptual frameworks for the consideration of natural phenomena[1–3]. For modern science, reading medieval works presents several significant challenges, starting not least with that of editions and translations. This analysis of the *DGS* has only been possible through interdisciplinary collaboration between science and humanities scholars, resulting in the compilation of a new critical edition and translation of the text[45].

Previous interdisciplinary research has already explored other scientific treatises written by Grosseteste: the *De colore* [On Colour][6], *De iride* [On the Rainbow][7,8], and *De luce* [On Light][9]. In the *De colore*, Grosseteste develops a pioneering application of mathematics to psychology. Within the space of approximately 400 words, he claims that colour occupies a continuous, three-dimensional space, contrary to the prevailing one-dimensional theory of the time[6]. It is surprising to find this theory articulated six centuries before three-colour printing techniques were established[10] and trichomacy was formulated by Thomas Young[11]. In the *DGS*, the treatise we explore and respond to in this paper, Grosseteste attempts a similarly mathematical, combinatorial abstraction for phonetics—specifically for vowels—as he attempts for colour. Several features of how he goes about doing this are of interest to the modern reader. Whereas Grosseteste's colour space is explicitly continuous, the vowel space described in the *DGS* is explicitly categorical. Underpinning his theory is a multimodal framework identifying correspondences between the mental representation of vowels, their physical production, their perception, and their external representation as letter shapes. Within this framework, the correspondences between speech perception, letter perception, and shape perception, have particular modern resonances in audiovisual speech and involvement of the motor system during speech perception. In the second half of this paper we present an experimental interpretation of the text, using artificial vowel synthesis and psychophysics to test the claims of correspondence between abstract, geometric acoustic chamber shapes and vowel perception.

Before presenting a detailed discussion of the *DGS*, a question that might first be addressed is why one ought to concern themselves with medieval science. Modern neuroscience is already at an interdisciplinary juncture between psychology, physiology, biology and mathematics; why should matters be further complicated with the inclusion of medieval history and Latin? An answer may be found in the sheer wealth of scientific theory and observation that was amassed during this period, which largely remains untapped. The history of science is highly non-linear, despite its frequently linear presentation, leaving worthwhile questions and suggestions unresolved in every historical age[12]. Psychological phenomena such as the perception of speech are not new, and have been prompting rational discourse throughout many historical and geographical cultures. By engaging with these theories today, we may find unexpected agreement with, or perspectives that are strikingly different to, our own. In either case, we stand to gain much from the exercise.

II. ROBERT GROSSETESTE'S *DE GENERATIONE SONORUM*

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

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

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

The remainder of the treatise is an attempt to characterize those 'mental images' that initiate the voice, and the relationships between mental representations of origination, the physical gestures of the vocal tract, the acoustic qualities of vowels, and the movements of the hand that draw out letters to represent speech sounds[13]. Immediately following on from the above passage, Grosseteste demarcates the difference between an intelligible and an unintelligible speech sound:

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

Here, Grosseteste establishes a direct relationship between the shapes—or as they may be understood, *figures*—of mental images, vocal tract shapes, and the movements of the breath during speech. These three figures, when perfected, give rise to a 'lettered voice', i.e. an acoustic output of intelligible speech. Grosseteste does not yet describe these figures geometrically, though that will come in the next section of the treatise. It is interesting to note the particular emphasis on the natures of certain voices due to the ‘actualising shaping itself of the vocal instruments’; any voice is preceded by a mental image, but the intelligibility of that voice additionally depends on the speaker’s ability to precisely execute the required motor programs. Or, to further unpack this notion, the acquisition of speech requires first the presence of mental representations for speech sounds (it is unclear whether Grosseteste is of the opinion these are innate or acquired), and second the learning of distinct motor programs encoding muscular coordinations for the production of these speech sounds. While Grosseteste does not explicitly describe this in terms of language acquisition, and the development from an imperfect to perfected voice, it is heavily implied when understood in the broader medieval context of discussions on the liberal arts. The seven liberal arts—and in this case the first art, that of grammar—provide a means whereby the fallen and corruptible things of the world may be refined and perfected through study and practice. In this case, the notion of a ‘perfect’ or comp[illegible]ed voice is related to the art, and study, of grammar, and the acquisition of vocal tract coordinations that give rise to a ‘lettered’ voice, i.e. intelligible speech[14].

In isolation, it may seem from this passage that Grosseteste understands that both diaphragmatic breath control (‘shaping of the movement of breaths’) and muscular coordination of articulators (‘shaping of the vocal instruments’) are required to produce intelligible speech sounds. However, he later makes clear that he is instead claiming a direct identity between control of the vocal apparatus and the resultant movements of the (‘motive’) breath, and it is these motive breath shapings that determine the ‘outward appearance and perfection’ of a voice. Writing six hundred years before Fourier and modern notions of frequency, resonance, and spectral analysis, this provided a sensible hypothesis for the causal relationship between the shape of the vocal tract and the acoustic qualities of the generated sound.

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

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

Here Grosseteste is guided by two Aristotelian principles: first, that 'art imitates nature', or mimesis; and second, that nature always acts in the best possible way. There is clear indication of his reading Aristotle's *De anima* [On the Soul][15], although Grosseteste does not reference Aristotle directly, as he does in some othe[illegible]ientific works[16]. These principles motivate one of the most central and clearly articulated claims of the treatise: 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 speech production which is summarized in Figure 1. This claim that 'representation by exterior shapes assimilated to interior will be better than otherwise' is particularly interesting, and has strong resonances with recently resurfacing theories of non-arbitrary representation, or 'iconicity'[17,18].

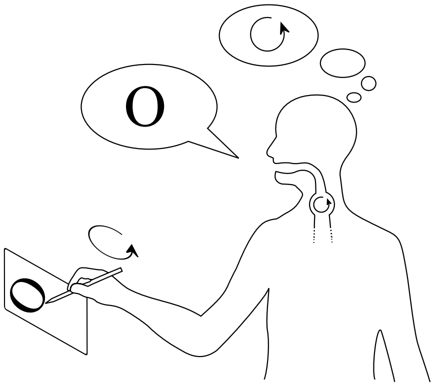

FIG. 1. A diagrammatic depiction of one of the claims in Grosseteste's *De generatione sonorum.* Grosseteste claims that the capacity for speech to be written lies in the visual representation of shapes similar to the geometric figures (mental, gestural, and of the 'motive breaths') at play during the 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.

For many languages today, including modern English, such a direct relationship between speech-sound (phoneme) and written letter (grapheme) would be impossible; individual letters have diverse pronunciations in differing lexical contexts, themselves quite different to the letter name. As an example of phonological inconsistency, while an English speaker with received pronunciation today may read the letter 'O' as a diphthong /əʊ/, it could be similarly pronounced as /əʊ/ in 'go', but also as /u/ in 'do', /ʌ/ in 'tonne', /ʊ/ in 'woman' and even /ɪ/ in ‘women’. This complication was not known to Grosseteste, who saw a mostly direct and consistent grapheme-phoneme relationship in the languages it is likely that he knew (Middle English, Latin, and French). Any exceptions, such as variations in regional accents, could be accounted for as being ‘accidental’.

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

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

In fact, this is a combinatorial system related to that described in the *De colore*: three simple elements are combined in various ways to give rise to a full set including complex combinations, except that for this scheme only two simple elements may be combined rather than all three. It is also different in that, rather than being defined by independent dimensions as in the case of the bipolar qualities of colours, only some of the simple elements may be combined, and one—circular movement—may be self-combined. The choice of three simple movements may not appear such an obvious choice, and it may be even more puzzling why only one of the three may be self-combined. Grosseteste states clearly that this is the comprehensive list of movements ‘in which the parts are the same as the whole’. We may rephrase this description as one of time-invariant functions on position.

One way of interpreting the scheme that seems to resolve these confusions is by viewing the three classes of simple movements as geometric linear transformations. In which case, these movements correspond perfectly to the allowed operations for Euclidean similarity transformations: straight movement for translation, circular movement for rotation, and dilational movement (and constrictional) as uniform scaling. Matrix notation provides a convenient and efficient way of describing these transformations; while Grosseteste would not have had this notation at his disposal, imagining these movements *per se* is not contingent on any particular form of mathematical description. Expressed as two-dimensional transformation matrices of translation, rotation, and scaling—$\mathbf{A_t}$, $\mathbf{A_r}$, and $\mathbf{A_s}$, respectively—these three simple geometric transformations are given as:

$$\text{Translation} : \mathbf{A_t} = \begin{bmatrix} 1 & 0 & t \\ 0 & 1 & t \\ 0 & 0 & 1 \end{bmatrix}; \quad \text{Rotation} : \mathbf{A_r} = \begin{bmatrix} \cos(t) & \sin(t) & 0 \\ -\sin(t) & \cos(t) & 0 \\ 0 & 0 & 1 \end{bmatrix};$$

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

Using this interpretive scheme, the geometric figures which Grosseteste describes naturally arise by the consideration of points in Euclidean space experiencing these transformations. These simple and combined movements may be visualized in Figure 2 and Figure 3, respectively, and in the videos included in the online version of this paper for translation (Video 1), rotation (Video 2), dilation and constriction (Video 3), rotation and translation (Video 4), and dilation/constriction and translation (Video 5).

Mm. 1. Translation. File of type mp4 (1.8 MB)

Mm. 2. Rotation. File of type mp4 (1.8 MB)

Mm. 3. Dilation and constriction. File of type mp4 (1.7 MB)

Mm. 4. Rotation and translation. File of type mp4 (1.8 MB)

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

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

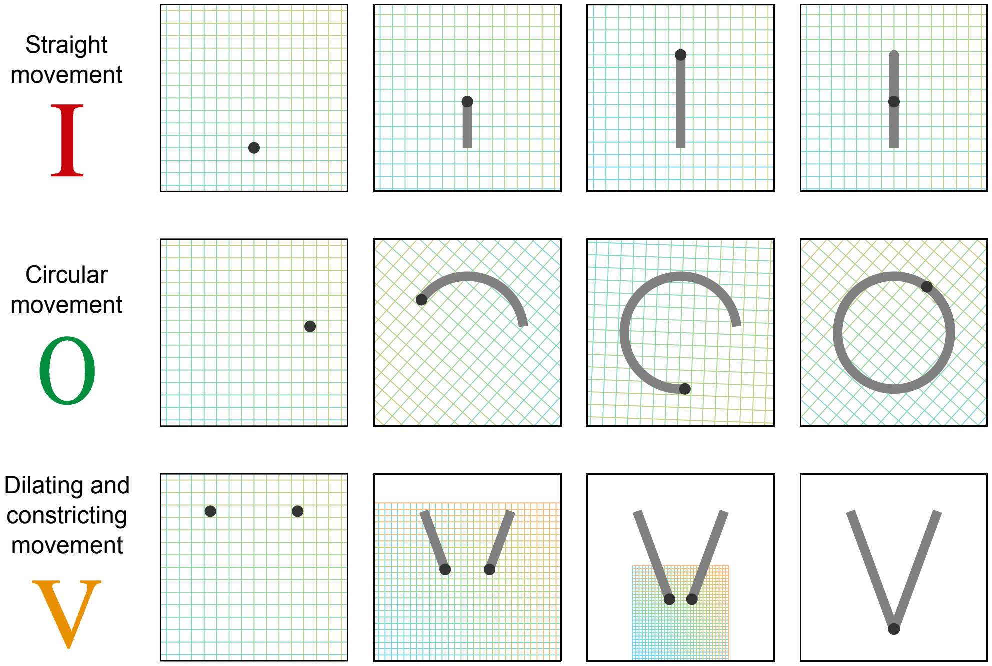

FIG. 2. The simple, self-similar geometric movements that Grosseteste describes as the basis for vowel categorization. We have interpreted his categories of simple movements—straight movement, circular movement, and dilating and constricting movement—as the three fundamental classes of linear geometric 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.

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

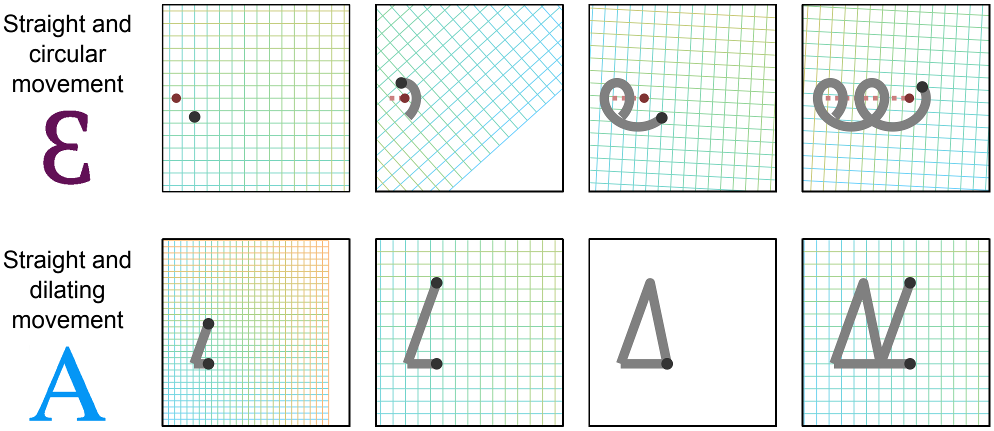

FIG. 3. The combined movements that give rise to vowels in Grosseteste's model of phonetics. For the combination of straight and circular movement, the translating origin of rotation is indicated by a red dot. For the combination of straight movement with dilating and constricting movement, two dots 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.

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

Mm. 6. Double rotation. File of type mp4 (1.8 MB)

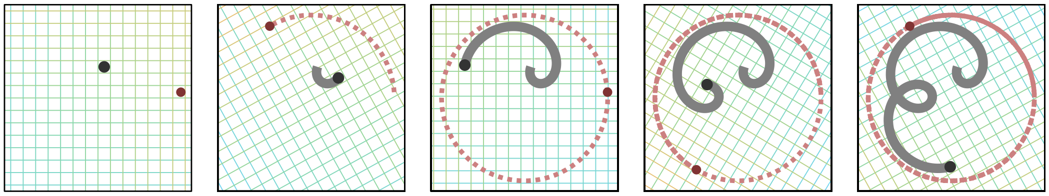

FIG. 4. Grosseteste describes a self-combination of circular movement, which he discounts as too complex for use in speech. This movement strongly evokes the mathematical constructions of epicycles in medieval astronomy. Here, the rotating origin of rotation is indicated with a red dot. Videos are provided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

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

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

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

> shape 'E'. A constricting and dilating movement, on the other hand, makes the figure 'V', that is, two lines running together in a centre. And a dilating and constricting movement over a centre moved straight in a straight movement subtends the base of a triangle. And any point, when there is dilation, because it is moved by a double movement, describes one side of the triangle from the base to the top, and when there is constriction, it describes the remaining side from the top to the base, and thus it makes the figure 'A'.

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

As made clear by these descriptions, the abstract figures that correspond to phonemes (and, on account of the art of grammar imitating nature, graphemes) are not static geometric shapes, but rather categories of movement, which are ascribed to the vocal tract during speech. Therefore, for Grosseteste the perception of a speech sound, whether in hearing speech or in reading, is intrinsically connected with vocal gestures, and the 'mental images' that encode their associated motor programs. This multisensory framework readily lends itself to current discussions of the motor theory of speech[19], and involvement of the motor cortex in speech perception.

Eight centuries after Grosseteste was writing, we now have experimental evidence from brain imaging and transcranial stimulation that his intuitions were solid. Involvement of the motor system was established fifteen years ago in response to visual and auditory speech perception[20], and soon after, that specific motor circuits in the precentral gyrus are recruited to facilitate phoneme identification—serving as 'speech-sound-specific neuronal substrates' shared across the sensory and motor processes[21]. Motor cortex involvement has been found to be beneficial for speech perception under noisy conditions[22], and possibly under normal listening conditions[23] (although possibly not[24]). Of particular relevance to Grosseteste's theory, Möttönen and Watkins (2009) found direct evidence for motor representations playing a complementary role in the categorization of speech sounds when they are found along continua[25]. As they point out, the mapping of highly variable acoustic signals onto discrete motor representations could support the intelligibility of speech in challenging environments. Even more intriguingly, Tian and Poeppel proposed a common sequential estimation mechanism underpinning both the quasi-perceptual experience of articulator movement and the corresponding auditory percept of speech mental imagery[26]. They claim that the experimental evidence from both task demands and stimulus properties demonstrates the top-down role the motor system is playing in this type of mental imagery. In which case, Grosseteste's claim that the mental imagery of speech is in fact a mental representation not of sound, but of motion (albeit of a simple, geometric nature), was remarkably apt.

In the light of these recent investigations, we can again consider Grosseteste's approach to understanding speech. Acoustic signals show enormous variety, and to the thirteenth-century researcher writing before the advent of spectral analysis, this would have proved impossible to organize. Confronted with the curse of dimensionality, Grosseteste limits his study of sound to that of speech—a subset of natural sounds that the human auditory system can reliably organize, doing so in a categorical manner. Aristotelian principles, the scientific paradigm of the day, provide the methodological approach, with the movements of the hand during writing perhaps constituting a permissible form of evidence for understanding the mental and anatomical origins of speech, and its perception. That speech sounds differ due to differences in movement category sits well with what Grosseteste understands about the vibrational mechanics of sound; sound is the perception of a special class of movements made by physical bodies, either when struck (the sounding body) or when formed by a primary motive force capable of forming mental images (the voice).

III. A PSYCHOPHYSICAL INTERPRETATION OF THE TEXT

The claims in the *DGS* are bold, and may read today as ‘unscientific’, lacking any evidential basis. But before dismissing these claims out of hand, it is worth considering exactly what evidence would have been available at the time to a shrewd observer. The morphology of the vocal tract would largely have been unknown, although from the end of the twelfth century very good diagrams of the vocal tract and its articulators were being produced in the Arabic-speaking world[27]. These would not have been accessible to Grosseteste, and we can reasonably say that any data he had regarding vocal tract morphology would have come from his own direct experience of vision and proprioception. As has been remarked by others, the resemblance of the ‘O’ letter shape and the pronounced rounding of the lips when producing the /ɔ/ phoneme may suggest a non-arbitrary grapheme-phoneme relationship[28], and could have been a motivating factor for the theory as a whole.

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

A. Stimuli

Synthetic vowels were produced by plate-type model vocal tracts, constructed to resemble the five geometric figures Grosseteste describes at the mouth end. This is a one-dimensional model developed by Arai *et al.*[29], comprising 75 mm wide acrylic squares, each 10 mm thick, with central holes of different diameters. The plates are clamped together in a specified order, leaving a central cavity of varying size down the length of the tract. A rubber coupler allows the introduction of an electrolarynx to acoustically stimulate the model at the laryngeal end, which produces a falling pitch excitation in the male range from 100 Hz to 60 Hz lasting around two seconds. Adjustments were made to the laryngeal end of the models such that the output best approximated the associated phoneme. The resultant plates are shown in Figure 5, which also includes an overlay in red of the region made to resemble the geometric shape for each vowel. The acoustic outputs of these vocal tract models were then analyzed acoustically (formant analysis) and perceptually (two psychophysical listening tests), to evaluate how successfully the synthetic speech-sounds approximate natural vowels.

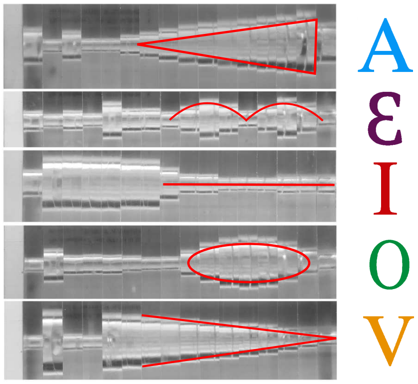

FIG. 5. The configurations of the plate-type vocal tract model (VMT-10) of Arai *et al.*[29] used to synthesize the five samples corresponding to Grosseteste's geometric figure associations for each of the five vowel letters, with the mouth-end on the right. From top to bottom: A, $\mathcal{E}$, I, O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

B. Formant analysis

Spectrograms for each sample were generated with a Hamming window of 20 ms, as shown in Figure 6, Upper Panel. The Lower Panel shows smoothed spectral slices calculated as the mean of each spectrogram across time. Difference between these synthesized stimuli and natural vowels are the shape of the acrylic plates vs the speaker's vocal tract—which is our primary interest—and the acoustic excitation (electrolarynx vs a speaker's larynx). The electrolarynx for the Arai tubes provides a signal that has a constant spectrum whereas the output from the vibrating vocal folds of the speaker vary as a function of the airflow loading owing to the shape of the vowel being uttered, sub-glottal lung air pressure through breath control and the nature of the voice quality being employed and any pitch variation.

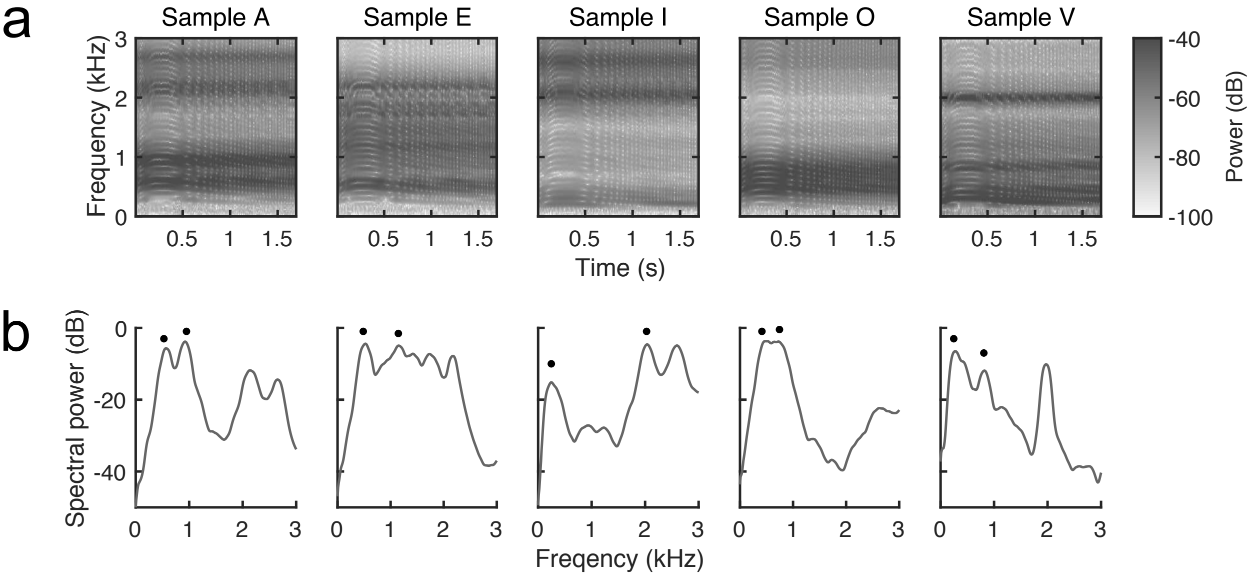

FIG. 6. (a) Spectrograms produced from each of the five synthesized samples. (b) Spectral slices given by the mean of each spectrogram across time for each sample, from which the frequencies of the first two formant peaks,F_1 and F_2, were taken (indicated by black dots).

The horizontal dark bands in the spectrograms show formants (peaks in spectral power) that result from filtering the input acoustic excitation of the electrolarynx by the passive acoustic resonances of the chambers. The primary acoustic features of vowels are the locations in frequency space of their two lowest-frequency formants, F_1 and F_2. When, for different vowels, F_1 is plotted on the ordinate and F_2 is plotted on the abscissa, the vowel quadrilateral results, and different vowels plot in well-separated regions of this acoustic space ([30] p. 161). A vowel quadrilateral for the synthetic vowels produced via the plate-type model is shown in Figure 7. This plot confirms that the acoustic properties of the synthetic samples are broadly consistent with the patterns of formants of natural vowels documented in the prior literature, with all samples falling within the quadrilateral. Additionally, the samples locate to disparate regions of the quadrilateral, suggesting they may be perceived as separable vowels.

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

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

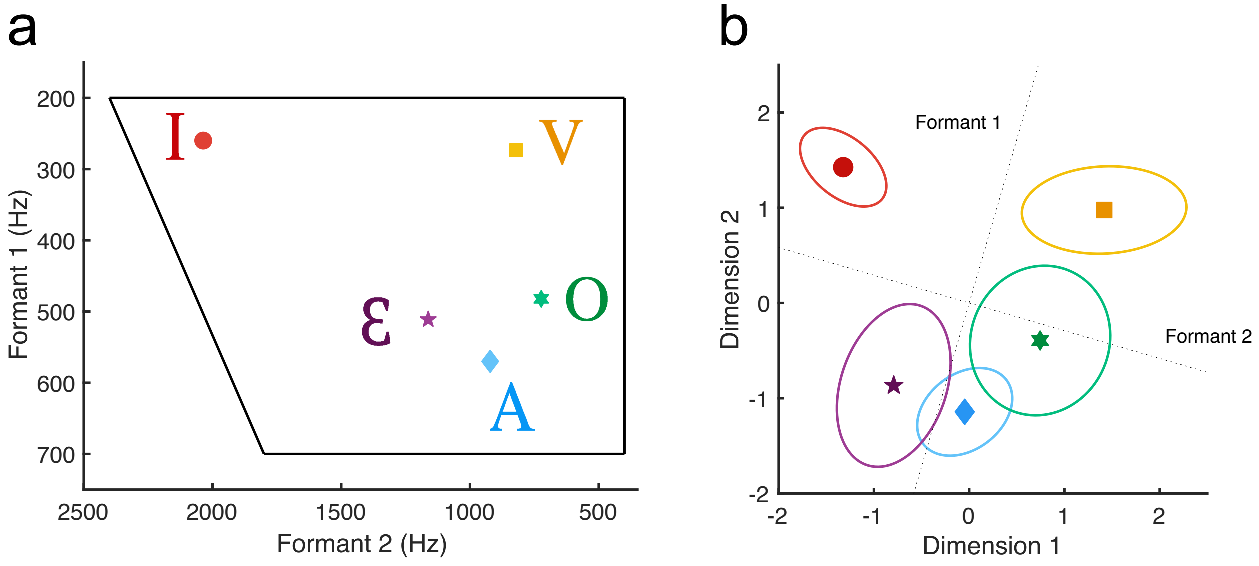

FIG. 7. (a) Acoustic map of the recorded synthetic vowels based on their measured first and second formant (F_1 and F_2) frequencies. The quadrilateral indicates the area within which discernible vowels are expected from previous literature[30]. Blue diamond = sample A, purple pentagram = sample $\mathcal{E}$, 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.

of vowel perception have broadly remained constant to the modern era, although some finer elements of speech perception vary as a result of differing cultural and language contexts[31]. For this reason, we selected participants from a range of language backgrounds.

C. Multidimensional scaling experiment

In the first psychophysical experiment the five stimuli were presented to both native and non-native English speakers to obtain dissimilarity scores. The *.wav* files (sampling rate 44,100 Hz, 16 bit, monophonic) were all normalized to 0 dB relative to full scale and limited to a duration of 1.70 s in Audacity, to be played through a pair of Sennheiser HD201 Closed Dynamic Stereo headphones. The experiment was built using the open-source Matlab function set Psychtoolbox[32], and run using the same laptop and headphones in quiet conditions. 20 participants took part in the experiment (12 female, 8 male, mean age 25 years). Participants were asked for their country of origin (13 UK, 1 USA, 2 India, 2 Bulgaria, 1 Germany, 1 Poland), if they were native or non-native English speakers and if non-native what their native language was (16 native English speakers [13 monolingual UK, 1 monolingual USA, 2 bilingual in English and Hindi], 4 non-native [2 Bulgarian, 1 German, 1 Polish]).

Participants were first played each of the five stimuli once for familiarity. Pairs of recordings were then presented separated by a 300 ms pause, and participants registered their perceived dissimilarity via a keyboard, from 0 (identical) to 7 (very dissimilar). For stimuli $i, j = 1, \ldots, 5$, all possible pairs were presented once in a random order, for both (i, j) and (j, i) sequences, to give a dissimilarity response matrix. From this, a symmetric matrix was constructed for each participant by taking means of (i, j) and (j, i) values. For six of the participants a single set of dissimilarity judgments was collected, while 14 went through the experiment twice. Since no systematic differences in dissimilarity scores were found between repeats, their symmetric matrices were averaged.

Kruskal's non-metric multidimensional scaling (MDS)[33] was performed on the symmetrical matrices to approximate the relative locations in perceptual space of the samples for individual participants. Once Euclidean coordinates were obtained from MDS analysis, these were plotted to inspect their agreement with the formant plots of the samples. Visual inspection of the mappings showed a clear correspondence between the first dimension of scaling and F_2, and the second dimension of scaling and F_1, for the majority of participants, which was later formally analyzed as described below. This agrees with previous studies that find human vowel discrimination primarily tracks the frequency position of F_2, which corresponds to perceived vowel advancement, and secondarily tracks the frequency position of F_1, corresponding to perceived vowel height[34]. There were four exceptions for this agreement; notably, these data sets were from the four non-native English speaking participants. Further inspection showed that these data agreed with F_2 and F_1 when plotted in the first and third dimension from the MDS, respectively, and hence these mappings were taken forward in the analysis.

Data sets then underwent Procrustes analysis, which permitted similarity transformations of the mappings (uniform scaling, orthogonal rotation, translation and reflection) in order to give the best concordance across participants while maintaining relative perceptual distances within mappings[35]. Once realigned, data sets were analyzed to extract the statistics for each stimulus as located in perceptual space by participants. Figure 7.b shows the mean positions for each stimulus, plotted as solid symbols. Ellipses show one standard deviation of the bivariate distribution of each vowel within the two dimensions of scaling. Sample O gave rise to the most spread compared to the other vowels, indicating that participants differed most in where to locate it in their perceptual space, relative to the other vowels. This is likely related to the strong degree of variation present in open back vowel pronunciations across dialects of English.

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

Monte Carlo simulations were carried out to evaluate the likelihood of stimuli being mapped to distinct regions due to chance, and consistently with the same relative orientation. From 26 simulations, only 20 generated data that could be mapped by MDS. After Procrustes analysis of these 20 mappings, none gave rise to a distinct region for any of the stimuli (i.e. non-overlapping regions bound by one standard deviation of stimuli mean position), and all stimuli regions had an area above 5 scaling space units^2, compared to a mean of 1.2 scaling space units^2 for participant-generated data. For all mappings, shown in Figure 9 in the Supplementary Material, the relative orientation of vowels were different. A more extensive simulation was carried out to generate 100 mappings, whose ellipses had a mean of 7 scaling space units^2, shown in Figure 10. We therefore conclude that the results of mapping the participant data, with stimuli occupying separable regions and a relative orientation in agreement with the acoustic analysis, are not owing to chance.

D. Vowel classification experiment

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

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

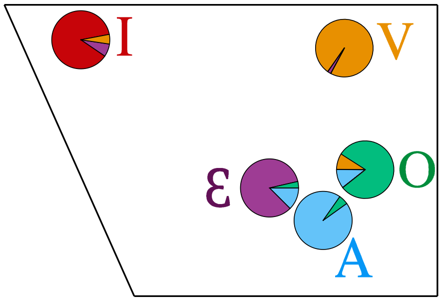

FIG. 8. Classifications obtained for each of the five samples from the second listening test. The pie charts for each sample, showing participants' classifications, are centered at the samples' locations when mapped in acoustic space, as shown in Figure 7 a. Responses are indicated by color: "'ah' as 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.

E. Results: MDS and classification experiments

Listening to isolated vowels is not a common activity in daily life, and listening to isolated vowels without having any reference to the speaker is also unusual. In addition, these stimuli are clearly non-human in origin given the identical electrolarynx acoustic input in each case. Some confusion is therefore inevitable. As may be expected, the synthetic vowel with the broadest spread of placement in perceptual space (indicated by its ellipse in Figure 7 b having the greatest area) was also the least reliably classified sound, Sample O, which received 80.4% correct classifications and 10.7% and 8.9% misclassifications as 'ah' and as 'oo' respectively. The greatest source of misclassification was the assigning of both Sample $\mathcal{E}$ and Sample O as 'ah' (12.5% and 10.7% respectively). The perceptual space generated by MDS analysis and the acoustic space from formant data both show Sample E and Sample O located in close proximity to Sample A, which itself was classified as 'ah' with high agreement. Indeed, on the perceptual map these are the only two instances of overlapping standard deviations from the samples' means. It can be said with confidence that the samples are perceived, imperfectly, as vowels, spanning a large proportion of vowel perceptual space.

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

IV. DISCUSSION

While sometimes described as a scientist, and undoubtedly instrumental in the conception of the scientific experimental method[36], we must be careful when reading Grosseteste's treatises not to impute any sense of experimental or even observational basis for his theories, however elegant the logical or mathematical arguments found therein. Recent interdisciplinary research has found that the origin of such theories, though they may be wrong within the context of current scientific understanding, may still best be explained as resulting from direct observation, such as for his novel theory of rainbow formation[8]. However others, though they may have been correct, are unlikely to have had a direct observational basis, such as his three-dimensional theory of colour space as expressed in the *De colore*[6]. These works remain remarkable achievements, and the desire to mathematicize the mental or material world was a fundamental evolution for intellectual history in the medieval and early modern era.

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

The scheme is defined by what he says are the three types of simple, self-similar movements: linear, circular, and dilational-constrictional. These simple movements may be combined, but only a subset yield novel categories of movement: combining linear with circular, linear with dilational-constrictional, circular with circular, and circular with dilational-constrictional. These descriptions of movement are readily interpreted as the three types of geometric similarity transformations—translation, rotation, and uniform scaling (with reflection being equivalent to rotation through a higher dimension)—though it should be noted that no diagrams are found in extant manuscripts, and this is just one possible interpretive scheme[37]. The treatise can be read as one primarily about types of movement, and relies heavily on the false premise that sounds of different qualities are discriminable based on the category of vibrational movement, rather than the spectral filtering achieved by differently-shaped acoustic chambers with varying resonant frequencies, and other language-specific factors. Although this theory is mistaken about the underlying source of vowel timbre, Grosseteste nevertheless constructs an elegant theory that attempts to account for the categorical nature of vowel perception, and the representation of vowels as letters.

Reading this text today prompts us to examine what may constitute permissible evidence in science. For Grosseteste, the shapes of letters could serve as the primary evidence for his claims regarding the shape of the vocal tract, and the forms of mental representations of vowels; within the medieval paradigms of Aristotelian mimesis and the liberal arts, this was a scientifically orthodox and justifiable use of observations to infer properties of the natural world. Although we do not share these paradigms as modern scientists, we share in the methodological framework of setting our own standards for permissible evidence; in many cases such sources of evidence are far-removed from the phenomenon we attempt to study. A generous reading of the *DGS* could be that Grosseteste is engaged in modelling; do abstract movement categories offer a viable framework for the robust, categorical representation and perception of speech sounds, despite their continuous variety and noisy instances? Although our models of speech processing have matured in their awareness of acoustics and physiology[38–40], they share the underlying goal of understanding how speech signals are processed and represented.

The *DGS* does make strong claims about the morphology of the vocal tract during vowel production, which are clearly incorrect in asserting the presence of geometric shapes. However, we have shown, through artificial vowel synthesis and the methods of spectral analysis and psychophysical testing of vowel perception, that these geometric shapes can in fact be incorporated at the mouth end of acoustic chambers that give rise to discriminable vowel sounds. This is plausibly due to degree of freedom present in the remainder of the acoustic chamber, i.e. the laryngeal and pharyngeal cavity, and the many-to-one property of acoustic chambers and their spectral output[41], meaning that unique speech sounds may have multimodal or highly nonlinear mappings in articulator space[42]. In the thirteenth-century Grosseteste would only have had visual and proprioceptive measurements of the lips, teeth and tongue, so any requirements of the rest of the vocal tract for vowel production could not have impacted his theory.

How influential the *DGS* was on the developing field of phonetics is difficult to say. Roger Bacon, a student of Grosseteste's who praised his mathematical approach to understanding nature, describes similar notions of relating the number of vowels in languages to the number of fundamental classes of movements in his text on Greek Grammar[43]. However, he seems to criticize these theories as falling outside the scope of the 'pure grammarian', instead they should be left to the disciplines of metaphysics and of music[44]. Specifically, he is engaging with the content of the *Tractatus de grammatica*. Circulating at the time, the anonymous *Tractatus* was widely attributed to Aristotle, but Bacon shows this to be unjustified, and the treatise was later sometimes ascribed to Grosseteste.

Readers familiar with Hangul, the native Korean alphabet devised by King Sejong the Great (1397-1450) in the fifteenth century, may find similarities between Grosseteste's theory of non-arbitrary letter shapes and the apparent similarity between Hangul consonant forms and their corresponding places of articulation[45]. However, we have no record of a reception of Grosseteste's work in east Asia, and any direct connection seems improbable. Moreover, while the articulatory basis of the Hangul alphabet is often stated as matter of fact, and has been written about since only a few years after Hangul was devised (such as in *Hwunmin cengum haylyey* [Explanations and Examples of the Correct Sounds for the Instruction of the People], published in 1446), there are competing theories. It seems equally likely that Hangul consonants were instead influenced by or modelled on the Mongol 'Phags-pa alphabet, itself derived from Tibetan, as suggested by Keith Whinnom[46]. It could therefore be the case that in Hangul and its reception we find a thesis parallel to claims made in the *DGS*: the notion of glyph iconicity being used as a kind of pedagogical or philosophical device to explain their forms.

Theories attempting to draw direct relationships between the shaping of articulators and the shapes of letters surfaced again in the seventeenth century, with Franciscus Mercurius van Helmont claiming that intrinsic to the Hebrew alphabet was found a phonetic guide to its pronunciation[47], and Bishop John Wilkins attempting to construct a visual alphabet of speech sound diagrams[48]. In neither case is there an explicit connection to the *DGS*. Such theories relating letter shapes to vocal tract shapes paved the way for the speaking machine of Wolfgang von Kempelen in 1780, and, later, the set of 'visible speech' symbols by Alexander Melville Bell[49,50].

Lastly, an essay published in 1772 by Charles Davy makes near identical claims regarding the representations of the vocal tract in the letter shapes of vowels[51] (p84-87), but again, any connection to Grosseteste's theory is not made explicit and may be entirely accidental. It should also be noted that Davy's text was not written as a serious scientific endeavour but as an amusing romp through classical trivia, with Davy himself writing: "The Editor will not undertake to defend it: as a *whimsical* conjecture, it may still afford some entertainment. *Better* reasons might perhaps be offered in its favour than what appear at present", before stating his belief that the Greeks' visual representation of the vocal tract in letter shapes is what enabled their literary success. It may simply be the case that such theories were best appreciated as a form of intellectual entertainment, rather than serious scientific endeavour. Now, with the advent of recent studies into glyph iconcity[17,18], theories of non-arbitrary representation of letter shapes are again being considered, albeit from a more nuanced and experimental standpoint.

V. CONCLUSION

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

ACKNOWLEDGMENTS

This work was supported by the AHRC under grant AH/N001222/1. JSH holds a DPhil studentship from the Andrew W. Mellon Foundation, hosted by the Oxford interdisciplinary research center TORCH. HES was supported by a Visiting Fellowship from the Institute of Advanced Study at Durham University. The work presented here emerged from the collaborative Ordered Universe Project (http://ordered-universe.com/), which focuses on interdisciplinary readings of the scientific works of Robert Grosseteste (c.1170–1253).

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

APPENDIX: SUPPLEMENTARY MATERIAL

Letter shape	Phoneme	Example
A	/ɑ/	‘ah’ as in ‘part’
$\mathcal{E}$	/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.

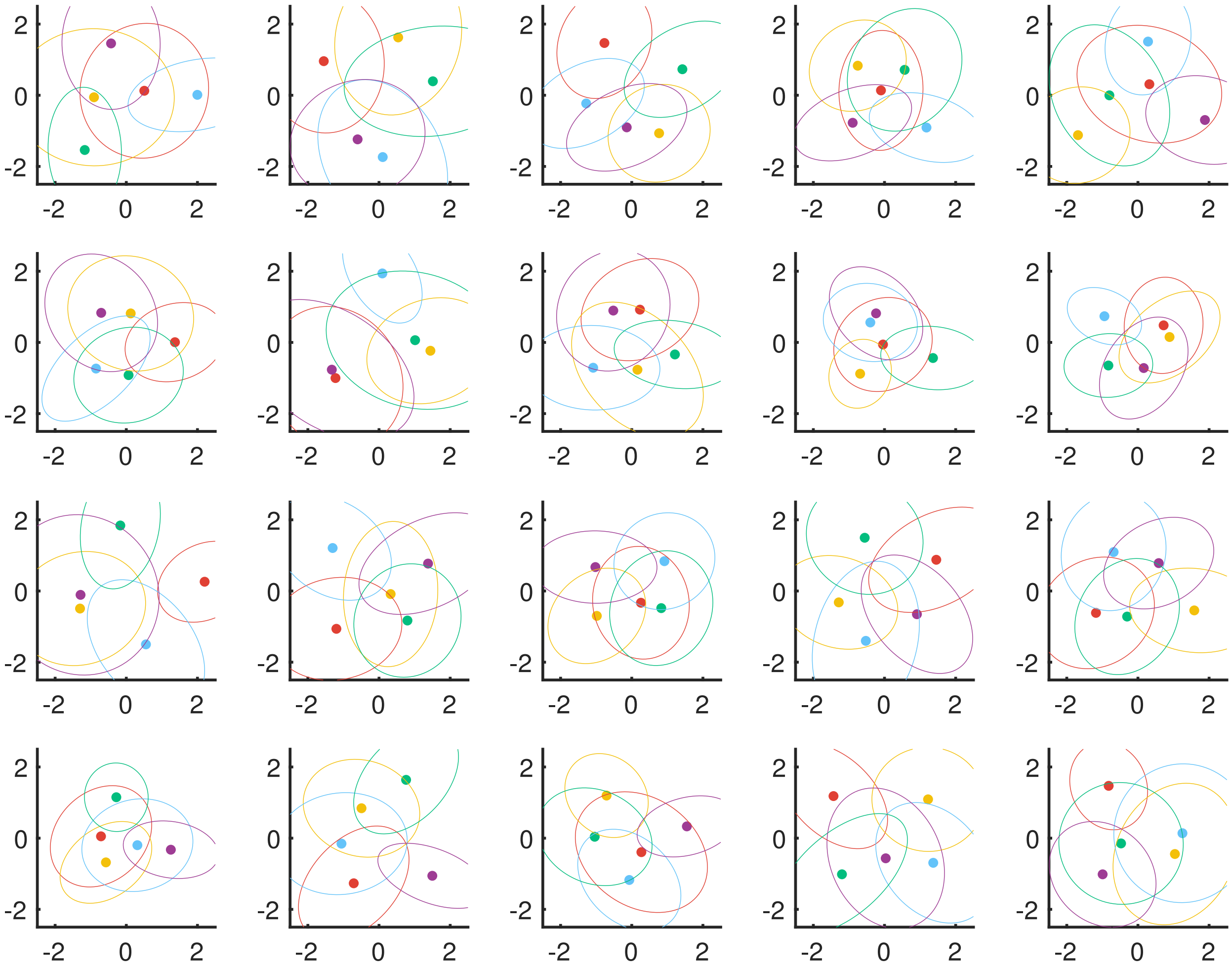

FIG. 9. 20 examples of Monte Carlo simulations that generated data sets for which a MDS mapping was possible. No simulation produced dissimilarity data that when mapped featured a distinct area for a stimulus, as bound by one standard deviation from its mean position (indicated by ellipses).

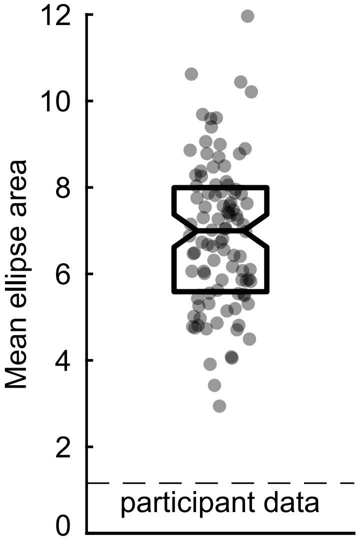

FIG. 10. The results of 100 Monte Carlo simulations of the MDS experiment. The mean ellipse areas from each simulation (which comprised 20 randomized participant data sets) are shown. The box plot indicates the mean and quartiles of the distribution, with a 95% confidence interval on the mean shown as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

References

[1] C. Burnett, D. C. Lindberg, and M. H. Shank, “Translation and Transmission of Greek and Islamic Science to Latin Christendom,” in *The Cambridge History of Science*, Vol. 2 of *The Cambridge History of Science* (Cambridge University Press, 2013), pp. 341–364, doi: `10.1017/CHO9780511974007.016`.

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

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

[4] This new critical edition and translation incorporates three manuscripts not known by the last editor, Baur[52].

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

by S. O. Sønnesyn.

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

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

[8] J. S. Harvey, H. E. Smithson, C. R. Siviour, G. E. M. Gasper, S. O. Sønnesyn, B. K. Tanner, and T. C. B. McLeish, "Bow-shaped caustics from conical prisms: a 13th-century account of rainbow formation from Robert Grossetestes De iride," Applied Optics **56**(19), G197 (2017) https://www.osapublishing.org/abstract.cfm?URI=ao-56-19-G197, doi: 10.1364/AO.56.00G197.

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

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

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

[12] H. Chang, "Who cares about the history of science?," Notes and Records of the Royal Society of London **71**(1), 91–107 (2017) `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5311899/`, doi: `10.1098/rsnr.2016.0042`.

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

[14] A comprehensive discussion of the seven liberal arts, and Grosseteste's treatise on them, may be found in[5].

[15] Aristotle, *Complete Works of Aristotle, Volume 1: The Revised Oxford Translation: Revised Oxford Translation v. 1* (Princeton University Press, Princeton, NJ, 1984), ed. Barnes, J.

[16] A deeper dive into the intellectual influences for Grosseteste's *DGS* can be found in[53].

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

[18] N. Turoman and S. J. Styles, "Glyph guessing for oo and ee: spatial frequency information in sound symbolic matching for ancient and unfamiliar scripts," Royal Society Open Science **4**(9), 170882 (2017) `http://rsos.royalsocietypublishing.org/lookup/doi/10.1098/rsos.170882`, doi: `10.1098/rsos.170882`.

[19] B. Galantucci, C. A. Fowler, and M. T. Turvey, "The motor theory of speech perception reviewed," Psychonomic bulletin & review **13**(3), 361–377 (2006) `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2746041/`.

[20] K. Watkins, A. Strafella, and T. Paus, "Seeing and hearing speech excites the motor system involved in speech production," Neuropsychologia **41**(8), 989–994 (2003) `http://linkinghub.elsevier.com/retrieve/pii/S0028393202003160`, doi: `10.1016/S0028-3932(02)00316-0`.

[21] F. Pulvermuller, M. Huss, F. Kherif, F. Moscoso del Prado Martin, O. Hauk, and Y. Shtyrov, "Motor cortex maps articulatory features of speech sounds," Proceedings of the National Academy of Sciences **103**(20), 7865–7870 (2006), doi: `10.1073/pnas.0509989103`.

[22] I. G. Meister, S. M. Wilson, C. Deblieck, A. D. Wu, and M. Iacoboni, "The Essential Role of Premotor Cortex in Speech Perception," Current Biology **17**(19), 1692–1696 (2007) `http://linkinghub.elsevier.com/retrieve/pii/S0960982207019690`, doi: `10.1016/j.cub.2007.08.064`.

[23] M. Sato, P. Tremblay, and V. L. Gracco, "A mediating role of the premotor cortex in phoneme segmentation," Brain and Language **111**(1), 1–7 (2009) `http://linkinghub.elsevier.com/retrieve/pii/S0093934X09000418`, doi: `10.1016/j.bandl.2009.03.`

002.

[24]A. DAusilio, I. Bufalari, P. Salmas, and L. Fadiga, "The role of the motor system in discriminating normal and degraded speech sounds," Cortex **48**(7), 882–887 (2012) `https://linkinghub.elsevier.com/retrieve/pii/S0010945211001614`, doi: `10.1016/j.cortex.2011.05.017`.

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

[26]X. Tian and D. Poeppel, "Mental imagery of speech: linking motor and perceptual systems through internal simulation and estimation," Frontiers in Human Neuroscience **6** (2012) `https://www.frontiersin.org/articles/10.3389/fnhum.2012.00314/full` doi: `10.3389/fnhum.2012.00314`.

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

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

[29]T. Arai, N. Usuki, and Y. Murahara, "Prototype of a vocal-tract model for vowel production designed for education in speech science.," in *INTERSPEECH* (2001), pp. 2791–2794.

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

[31]P. K. Kuhl, S. Kiritani, T. Deguchi, A. Hayashi, E. B. Stevens, C. D. Dugger, and P. Iverson, "Effects of language experience on speech perception: American and Japanese infants perception of /ra/ and /la/," The Journal of the Acoustical Society of America **102**(5), 3135–3136 (1997) `http://asa.scitation.org/doi/abs/10.1121/1.420646`, doi: `10.1121/1.420646`.

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

[33]J. B. Kruskal, "Nonmetric multidimensional scaling: A numerical method," Psychometrika **29**(2), 115–129 (1964) `https://link.springer.com/article/10.1007/BF02289694`, doi: `10.1007/BF02289694`.

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

[35]R. Sibson, "Studies in the Robustness of Multidimensional Scaling: Procrustes Statistics," Journal of the Royal Statistical Society. Series B (Methodological) **40**(2), 234–238 (1978) `http://www.jstor.org/stable/2984761`.

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

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

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

[39] K. N. Stevens, “Toward a model for lexical access based on acoustic landmarks and distinctive features,” The Journal of the Acoustical Society of America **111**(4), 1872–1891 (2002) http://asa.scitation.org/doi/10.1121/1.1458026, doi: 10.1121/1.1458026.

[40] G. Hickok and D. Poeppel, “The cortical organization of speech processing,” Nature Reviews Neuroscience **8**(5), 393–402 (2007) http://www.nature.com/articles/nrn2113, doi: 10.1038/nrn2113.

[41] B. S. Atal, J. J. Chang, M. V. Mathews, and J. W. Tukey, “Inversion of articulatorytoacoustic transformation in the vocal tract by a computersorting technique,” The Journal of the Acoustical Society of America **63**(5), 1535–1555 (1978) https://asa.scitation.org/doi/10.1121/1.381848 doi: 10.1121/1.381848.

[42] C. Qin and M. . Carreira-Perpin, “The geometry of the articulatory region that produces a speech sound,” in *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers* (2009), pp. 1742–1746, doi: 10.1109/ACSSC.2009.5469741.

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

[44] I. Roser-Catach, “Roger Bacon and Grammar,” in *Roger Bacon and the Sciences: Commemorative Essays*, edited by J. Hackett (BRILL, 1997), google-Books-ID: Gy3Vp7TurVUC.

[45] Y.-K. Kim-Renaud, *The Korean Alphabet: Its History and Structure* (University of Hawaii Press, 1997), google-Books-ID: nonRl2cerIgC.

[46] E. R. Hope, “Letter Shapes in Korean Önmun and Mongol hPhagspa Alphabets,” Oriens **10**(1), 150–159 (1957) https://www.jstor.org/stable/1578766, doi: 10.2307/1578766.

[47] F. M. van Helmont, *Alphabeti veri naturalis hebraici brevissima delineatio* (1667).

[48] John Wilkins, *An Essay towards a Real Character and a Philosophical Language* (Samuel Gellibrand, London, 1668), http://archive.org/details/AnEssayTowardsARealCharacterAndAPhilosophicalLanguage.

[49] A. M. Bell, *Visible speech: the science of universal alphabetics; or Self-interpreting physiological letters, for the writing of all languages in one alphabet* (London, Simpkin, Marshall & Co., 1867), http://archive.org/details/visiblespeechsci00bellrich.

[50] H. Dudley and T. H. Tarnoczy, “The Speaking Machine of Wolfgang von Kempelen,” The Journal of the Acoustical Society of America **22**(2), 151–166 (1950) https://asa.scitation.org/doi/abs/10.1121/1.1906583 doi: 10.1121/1.1906583.

[51] C. Davy, *Conjectural observations on the origin and progress of alphabetic writing.* (T. Wright, London, 1772), https//catalog.hathitrust.org/Record/001181056.

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

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

[54] J. S. Harvey, R. C. White, H. E. Smithson, T. C. B. [illegible]Leish, D. Howard, and J. Coleman, "Instrumental motions: Shaping and perceiving speech sounds," in *Knowing and Speaking: Robert Grosseteste's De artibus liberalibus 'On the Liberal Arts' and De generatione sonorum 'On the Generation of Sounds'*, edited by G. E. M. Gasper, C. Panti, T. C. B. McLeish, and H. E. Smithson, Vol. 1 of *The Scientific Works of Robert Grosseteste* (Oxford University Press, Oxford, 2019), pp. 336–366.