*INTERSPEECH 2021*
30 August – 3 September, 2021, Brno, Czechia

# Model-based exploration of linking between vowel articulatory space and acoustic space

*Anqi Xu[1], Daniel van Niekerk[1], Branislav Gerazov[2], Paul Konstantin Krug[3], Santitham Prom-on[4], Peter Birkholz[3], Yi Xu[1]*

[1]Department of Speech Hearing and Phonetic Sciences, University College London, UK
[2]Faculty of Electrical Engineering and Information Technologies, UCMS, Skopje, RN Macedonia
[3]Institute of Acoustics and Speech Communication, Technische Universität Dresden, Germany
[4]Computer Engineering Department, King Mongkut's University of Technology Thonburi, Thailand

{a.xu.17, yi.xu}@ucl.ac.uk

## Abstract

While the acoustic vowel space has been extensively studied in previous research, little is known about the high-dimensional articulatory space of vowels. The articulatory imaging techniques are limited to tracking only a few key articulators, leaving the rest of the articulators unmonitored. In the present study, we attempted to develop a detailed articulatory space obtained by training a 3D articulatory synthesizer to learn eleven British English vowels. An analysis-by-synthesis strategy was used to acoustically optimize vocal tract parameters that represent twenty articulatory dimensions. The results show that tongue height and retraction, larynx location and lip roundness are the most perceptually distinctive articulatory dimensions. Yet, even for these dimensions, there is a fair amount of articulatory overlap between vowels, unlike the fine-grained acoustic space. This method opens up the possibility of using modelling to investigate the link between speech production and perception.

**Index Terms**: articulatory space, acoustic space, articulatory synthesis, British English

## 1. Introduction

Due to the well-known many-to-one mapping between articulation and acoustics [1], speech production is sometimes thought to be incommensurate with perception [2]. Yet speech units like vowels are after all produced by articulation, thus there is a need to understand the link between the two systems. The first systematic study of the acoustic vowel space was conducted by Peterson and Barney [3] on the formant analysis of ten American English vowels in /hVd/ context. Since then, numerous studies have reported formant spaces for vowels by plotting the 2D distribution of the first formant (F1) and the second formant (F2) for English [4] and many other languages [5]. There have been some theoretical discussions on the relationship between vowel production and perception. In the quantal theory [6], vowels in a given language are produced in certain manners that yield steady-state, distinct and robust acoustic regions. Likewise, Liljencrants & Lindblom [7] argued, in what is known as vowel dispersion theory, that vowel categories repel one another in the sound inventory. Diehl [8] proposed an auditory enhancement hypothesis that vowel articulation is coordinated to result in maximal auditory salience. Taken together, these theories share the assumption that vowel articulatory space is shaped for maximum separation in auditory space.

In contrast, much less research has been carried out with regard to the articulatory space. An early study by Johnson et al. [9] has demonstrated a segregated distribution of jaw height, lip position and tongue height in American English vowels based on X-ray data. Later on, the emergence of tongue imaging techniques such as electromagnetic articulography (EMA) and ultrasound have enabled detailed observation of vowel articulation. Ximenes et al. [10] have investigated the distribution of tongue dorsum position of Australian and American English vowels by examining EMA data as well as the corresponding acoustic space. Unlike the well-separated acoustic vowel space from an individual speaker, the tongue dorsum areas overlap, especially for mid vowels. It is suggested that the discrepancy between the normalized formant space and tongue dorsum space is on account of other unexamined articulators. Most articulatory studies are, however, limited to tracking the movement of a few articulators, while the role of other articulators such as the larynx and the tongue sides remain poorly understood[i].

More recently, there has been an increased interest in understanding the link between production and perception in speech. Dang [11] has utilized a non-linear dimensionality reduction method, Laplacian eigenmaps, to observe articulatory and auditory structure of English vowels. Whalen [12] conducted a principal component analysis (PCA) on the normalized X-ray data of fourteen articulatory dimensions and reduced the dimensionality to three. It was found that there is more variance in the three articulatory components than in the first three formants across the vowels for half of the speakers and vice versa for the rest. These attempts to reduce articulatory dimensions may have simplified the overall analysis of vowel articulation, yet much uncertainty remains about the exact role of each articulator in shaping the vowel auditory space.

The present study adopted a recently developed automatic articulatory synthesis technique [13]–[16], with the goal to explore the high-dimensional articulatory vowel space and to identify the articulators that contribute the most to the separation of vowels in the acoustic space. The method enables a thorough examination of the multidimensional articulatory

[i] MRI technology allows elaborate examination of speech production, but it is expensive and not widely available.



http://dx.doi.org/10.21437/Interspeech.2021-1422

space through analysis-by-synthesis. We trained a 3D vocal tract model [17] to learn eleven British English vowels in /hVd/ context guided by acoustic feature and a speech recognition system. Native listeners were asked to identify the learned synthetic words. The acoustic space and the learned articulatory parameters of the correctly identified vowels were analysed by both statistical analysis and classification.

# 2. Method

## 2.1. Material

Eleven target vowels, /i, ɪ, ɛ, æ, ɒ, ɔː, ʊ, uː, ʌ, ɑː, ɜː/, were embedded in real words, heed, hid, head, had, hod, hawed, hood, who'd, hudd, hard and herd. The reason for using this word set is threefold: 1) to be compatible with the classic literature, 2) to ensure that production and perception experiments can be carried out naturally by native speakers, and 3) to minimise the coarticulatory effect of onset consonants on the vowel. The recordings were made by a native male speaker of Southern British English in a quiet room.

## 2.2. Vocal tract model

The vocal tract model (Figure 1a) used is VocalTractLab 2.2 (www.vocaltractlab.de), a state-of-the-art articulatory synthesizer that calculates area functions for an acoustic simulation on the basis of a geometrical 3D vocal tract model [17]. The model is adapted from MRI data of a German male speaker, and is controlled by twenty vocal tract parameters, as shown in Table 1. The high dimensionality of this model allows us to perform a more comprehensive exploration of the vowel production process than previous research.

Table 1: *Vocal tract parameters in the model.*

| Parameter | Description |
|---|---|
| HX, HY | Horiz. and vert. hyoid positions JX, JA |
| LP, LD | Lip protrusion and vert. lip distance VS, VO |
| VS, VO | Velum shape and velum opening |
| TTX, TTY | Horiz. and vert. tongue tip positions |
| TBX, TBY | Horiz. and vert. tongue blade positions |
| TCX, TCY | Horiz. and vert. tongue body centre positions |
| TRX, TRY | Horiz. and vert. tongue root positions |
| TS1 – TS3 | Tongue side elevation from the anterior to the posterior part of the tongue |
| LMA | Lip minimal area |

## 2.3. Optimisation of vocal tract parameters

We developed an algorithm to train VocalTractLab to learn the eleven English vowels via analysis-by-synthesis, as illustrated in Figure 1. Specifically, the vocal tract configurations were iteratively adjusted until a best acoustic match with the target utterance was found. The optimization was done with simulated annealing [18], a stochastic algorithm that seeks an optimal solution through the coarse-to-fine criterion. Such an algorithm system can heuristically optimise models with many degrees of freedom, such as speech. The learning process started with a neutral position (schwa) followed by broad adjustments of the vocal tract parameters and gradually converged to a solution. Mel-frequency cepstral coefficients (MFCCs) were used as the acoustic features for training [19], which is widely used in speech recognition and machine learning based synthesis as a robust parametric representation of speech acoustics. The model was trained with 1000 iterations for each vowel. To obtain enough synthetic sounds and the corresponding vocal tract parameters, the learning process was run twenty times, and in each process the five synthetic vowels with the lowest acoustic errors were preserved, resulting in a hundred vowels for each vowel category.

After the vowels were learned, the whole words were synthesized by combining the vowel parameters with those of the initial consonant and the coda consonant from VocalTractLab. The duration of all the segments in the synthetic words resembles that of the original utterances. Intonation contours were also added to the words based on pitch targets learned from the target words using PENTAtrainer [20]. Because the learned vowels varied in synthetic quality, we passed all the synthetic words to the IBM Watson Speech-to-Text system with a language model of British English for recognition. For each word, we selected fifty of them with higher recognition confidence for use in the perceptual experiment[ii].

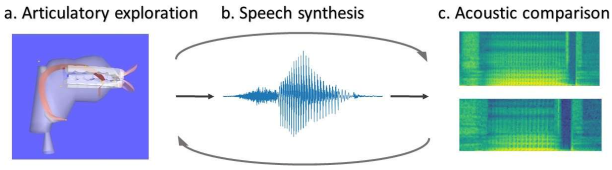


Figure 1: *Overview of the optimisation process.*

## 2.4. Perceptual experiment

Eleven Southern British English native speakers (female: 7; mean age: 33) participated in the online perception experiment. Fifty items of the eleven learned synthetic English vowels embedded in /hVd/ words (550 words in total) were randomised and presented to the subjects via Gorilla, an online experiment tool (gorilla.sc). In the experiment, participants were instructed to listen to the audio carefully and choose the word that they heard. There were twelve options including the word list and an additional option, 'none of the above'. They were allowed to listen to the audio up to three times. Listeners were asked to undertake the tasks on a computer in a quiet environment without noises or other distractions. A headphone check was conducted to make sure that the participants were wearing headphones and concentrated on the task. There were five practice trials preceding the listening test. The experiment lasted around 30 minutes.

## 2.5. Analysis

To visualize the acoustic vowel space, measurements of F1 and F2 of the steady-state vowels were extracted by FormantPro [21]. The articulatory vowel space was constructed from the optimized vocal tract parameters of the synthetic vowels selected by native listeners. Next, we conducted a non-parametric Mann-Whitney U-test to evaluate the difference in acoustic and articulatory space between all possible pairs of vowels. The significance level was 0.05. To examine the importance of each articulator in shaping the acoustic space, we trained decision trees, random forest and extra trees classifier to classify the words based on the vocal tract parameters. The more a vocal tract parameter contributes to the classification results, the more distinct it is across vowel categories. Extra

[ii] Audio samples can be found here: http://www.homepages.ucl.ac.uk/~uclyyix/EVL/hVd/

trees outperformed decision trees (75%) and random forest (85.14%) with a cross validation score of 85.27%. We trained extra trees for 10k iterations to obtain stabilised feature importance of the vocal tract parameters.

## 3. Results

Due to the nature of the online experiments, we established an inclusion threshold to ensure the quality of the responses. Three participants who failed to identify approximately half or more of the items were excluded from the analysis. We also excluded trials with reaction times less than the duration of the stimuli. Around 8% of the synthetic vowels remain unclassified ("none of the above") by the eight listeners included in the analysis (female: 6; mean age: 34). The identification rates of each vowel embedded in /hVd/ synthetic words are shown in Table 2, and the confusion matrix is shown in Figure 2. Fleiss's kappa [22] indicates that there is an excellent agreement for /heed/ (.78); a fair agreement for /had/, /hawed/, /head/ and /herd/ (.40 - .75); a poor agreement for /who'd/, /hudd/, /hood/, /hod/, /hid/ and /hard/ (< .40). The confusion was mainly between mid vowels. The learned /hudd/ was classified as /had/, /hid/ as /head/, and /who'd/ as /herd/.

The instances of eight vowels that have been identified by approximately half or more of the listeners were included in the analysis. As a result of the low consensus among the listeners, we did not include /hood/, /hudd/ and /who'd/ in the statistical analysis and classification, but the vocal tract parameters of the instances that have been identified by at least two listeners were plotted in Figure 4.

Table 2: *Identification rates of synthetic vowels embedded in /hVd/ words.*

| **heed** | **hid** | **head** | **had** | **hod** | **hawed** |
|---|---|---|---|---|---|
| 81% | 38% | 68% | 46% | 46% | 71% |
| **hood** | **who'd** | **hudd** | **hard** | **herd** | |
| 13% | 15% | 10% | 73% | 72% | |

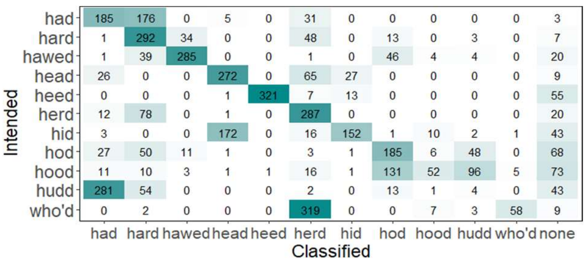


Figure 2: *Confusion matrix of the intended and classified vowels by eight native listeners*

The acoustic space of learned synthetic vowels and identified vowels are displayed in Figure 3. In general, the distribution is very similar but the regions of mid vowels are smaller after being classified by native listeners, especially for /hood/. Consistent with previous findings [10], the acoustic space of corners vowels in /heed/, /hawed/ and /who'd/ are distinct, whereas the areas for vowels in /hard/ vs. /hod/ are overlapped.

The distribution of the articulatory parameters for the identified vowel instances were plotted in a 2D articulatory space, as shown in Figure 4. Comparing Figure 3 and Figure 4, the variance in the articulatory space is much larger than in the acoustic space. The 2D space of the tongue body shows a clear

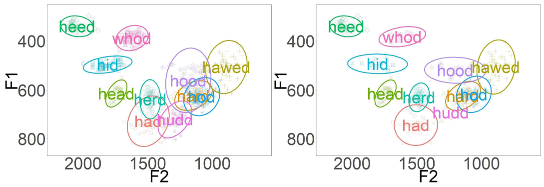


Figure 3: *Means and individual tokens of F1 and F2 for all learned vowels (left) and correctly identified vowels (right).*

grouping of vowel categories. In contrast to the acoustic space (Figure 3), the vowel areas in the articulatory space strongly overlap each other. If we take a closer look at the distributions of tongue body positions (TCX & TCY) in Figure 5, it is clear that TCX and TCY for mid vowels like /head/ and /hod/ are concentrated in a small region compared with the wide expansion for corner vowels. Rounded and unrounded vowels distributed separately into two groups in the lip protrusion and lip distance dimensions. The jaw angle parameter seems to be a reliable indication of vowel quality, showing that vowels also differ in the degree of jaw opening. Interestingly, vowels are distributed in relatively separated areas in the hyoid positions as well.

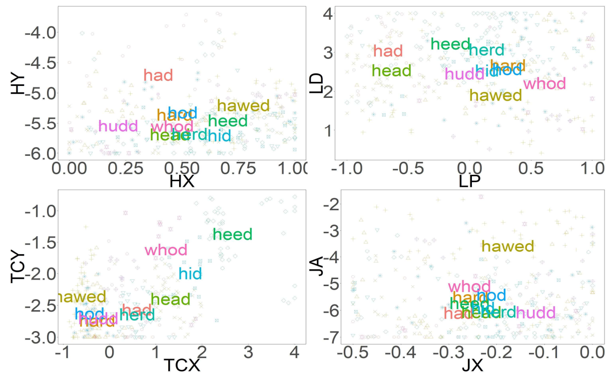


Figure 4: *Means and individual tokens of vocal tract parameters for learned vowels in the horizontal dimension and vertical dimension.*

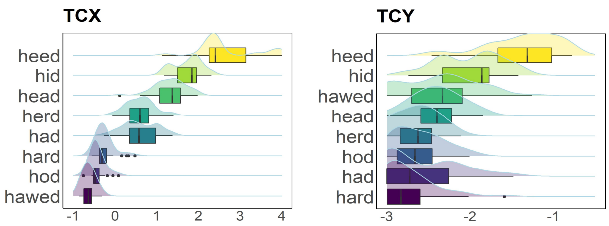


Figure 5: *Distribution of horizontal tongue body height position (TCX) and vertical tongue body position (TCY) for synthetic vowels*

To examine the variance in the acoustic and articulatory vowel space, we conducted statistical analysis on all the possible pairs of vowels (28 in total). A series of Mann-Whitney U-tests suggest that there are significant differences between the 23 pairs of vowels for F1 and 27 pairs of vowels for F2. The number of significant different vowel pairs in each articulatory dimension is summarised in Table 3. The difference between vowels is evident in the tongue body positions (TCX & TCY), tongue blade height (TBY), tongue side (TS2), hyoid positions (HX & HY), lip positions (LP & LD) and jaw angle (JA).

Tongue body locations in particular distinguish vowel pairs as efficiently as F1 and F2.

Table 3: *Number of significantly different vowel pairs in a Mann-Whitney U-test for 20 vocal tract parameters.*

| HX | LP | JX | VS | TTX | TBX | TCX | TRX |
|---|---|---|---|---|---|---|---|
| 14 | 19 | 5 | 7 | 7 | 4 | 27 | 3 |
| **HY** | **LD** | **JA** | **VO** | **TTY** | **TBY** | **TCY** | **TRY** |
| 20 | 17 | 13 | 7 | 6 | 12 | 23 | 6 |
| **TS1** | **TS2** | **TS3** | **LMA** | | | | |
| 0 | 10 | 3 | 9 | | | | |

In addition, we trained extra trees classifier to evaluate the importance of vocal tract parameters. Boxplots of the average feature importance values are shown in Figure 6. Consistent with the statistical results, tongue body parameters (TCX & TCY), jaw angle (JA), lips positions (LP & LD) and hyoid height (HY) contribute the most to the vowel quality. Tongue height and tongue retraction seem to be the most deterministic articulator dimensions that play a role in shaping the acoustics.

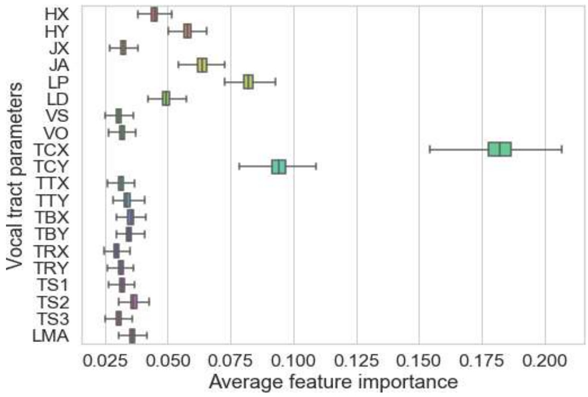


Figure 6: *Average feature importance of 20 vocal tract parameters by Extra Trees with 10k iterations*

## 4. Discussion

The current study investigated the relationship between the articulatory vowel space and the acoustic vowel space via high-dimensional articulatory synthesis. We trained a vocal tract model to learn eleven British English vowels in /hVd/ context and played the synthetic sounds to native listeners for identification. Our model performed reliably in finding appropriate vocal tract parameters of most of the vowels, and the acoustic space of the optimised vowels is similar to the one selected by native listeners, except for some mid vowels. Previous literature likewise reported that instances such as [ɪ], [ɛ], [ɒ] and [ɔ] produced by native American English speakers are often misunderstood [3], [4]. For the most extreme case, i.e., /ɒ/, less than 6% was classified unanimously [3]. It is worth noting that mid vowels in British English vary greatly with accent [23] while the recognition system that we used was not specifically designed for the Southern accent (i.e., the accent of the listeners). The recognition rate of /who'd/ was relatively low, even though the acoustic space of [u:] did not diverge much as verified by native listeners (Figure 3). /who'd/ learned by our model was classified as /herd/ because of the lack of lip protrusion, which is similar to the unrounded [u:] learned by the congenitally blind population [24]. Further work is required to incorporate visual input to the training model.

Despite the fine-grained acoustic vowel space (Figure 2), the analysis of the optimised vocal tract parameters indicates that there is no distinctive articulatory area among all the twenty dimensions (Figure 3). Consistent with previous research on the variance in articulation and acoustics of individual speakers [10], the distribution of articulatory parameters is much more variable than the corresponding acoustics. The follow-up statistical analysis further confirms that not all of the articulatory dimensions can sufficiently differentiate vowel pairs but tongue body positions perform similarly to F1 and F2. However, the data must be interpreted with caution because there are cases where some speakers demonstrate more variance in acoustics than articulation [12] and the results may vary according to the articulatory model.

The discrete perceptual space can nonetheless be inferred from the variable articulatory space in some dimensions better than others. The statistical analysis and classification results show that tongue body positions are reliable indicators of vowel quality, followed by vertical hyoid position and lip protrusion. As a main determiner of vowel quality, the tongue body positions for the same vowel indeed cluster but are not entirely separated between categories. The results corroborate the findings of [10] that the articulatory regions of the mid vowels overlap with one another. Lips and jaw locations were found to impact on the formants to a large extent, which supports the observations in [9]. Interestingly, we have also seen that the precise control of the hyoid position is indispensable to perceptual distinctiveness. In line with an MRI study on the movements of the larynx, the location of the hyoid bone plays an important role in vowel quality [25]. As suggested by [26], the tongue, the lips and the larynx function together on a reciprocal basis to produce [i] and [y]. This seems true for other vowels as well. The present study therefore provides further evidence that only when the speaker simultaneously coordinates several crucial articulators that the desired vowel can be precisely articulated, as argued in [27].

What is also interesting is that for mid vowels such as those in /head/ and /hod/, the distributions of tongue body height and frontness are more concentrated than other vowels. The limited perceptual space for mid vowels would demand particular articulator manoeuvres that involve many degrees of freedom. Developmental studies have shown that mid vowels are acquired later than corner vowels [28], which coincides with the lower learning performance for mid vowels in the present study. Despite the disadvantageous narrow acoustic space of mid vowels, learners ultimately manage to find a certain combination of articulatory postures to separate the restricted perceptual space into subregions. The current data support previous theoretical accounts of vowel perceptual space that the vowel system is auditory-based [6]–[8] rather than articulatory-based [29]. But articulation is equally important because only through trial and error as simulated by the analysis-by-synthesis process in the present study, can an equilibrium be reached between the speaker and the listeners, such that optimal articulatory manoeuvres can be eventually discovered.

## 5. Acknowledgements

This work has been funded by the Leverhulme Trust Research Project Grant RPG-2019-241: "High quality simulation of early vocal learning".

## 6. References


[1] B. S. Atal, J. J. Chang, M. V Mathews, and J. W. Tukey, "Inversion of articulatory-to-acoustic transformation in the vocal tract by a computer- sorting technique," *J. Acoust. Soc. Am.*, vol. 63, no. 5, pp. 1535–1555, 1978.

[2] C. Dromey, G. Jang, and K. Hollis, "Assessing correlations between lingual movements and formants," *Speech Commun.*, vol. 55, pp. 315–328, 2013.

[3] G. E. Peterson and H. L. Barney, "Control Methods Used in a Study of the Vowels," *J. Acoust. Soc. Am.*, vol. 24, no. 2, pp. 175–184, 1952.

[4] J. Hillenbrand, L. A. Getty, M. J. Clark, and K. Wheeler, "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.*, vol. 97, no. 5, pp. 3099–3111, 1995.

[5] A. R. Bradlow, "A comparative acoustic study of English and Spanish vowels," vol. 97, no. 3, pp. 1916–1924, 1995.

[6] K. N. Stevens, "On the quantal nature of speech: evidence from articulatory-accoustic data," in *Human Communication*, P. B. Denes and E. E. David Jr, Eds. New York: McGraw Hill, pp. 51–66, 1972.

[7] J. Liljencrants and B. Lindblom, "Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast," *Linguist. Soc. Am.*, vol. 48, no. 4, pp. 839–862, 1972.

[8] R. L. Diehl and K. R. Kluender, "On the objects of speech," *Ecol. Psychol.*, vol. 1, no. 2, pp. 121–144, 1998.

[9] K. Johnson, P. Ladefoged, and M. Lindau, "Individual differences in vowel production," *J. Acoust. Soc. Am.*, vol. 94, no. 2, pp. 701–714, 1993.

[10] A. B. Ximenes, J. Shaw, and C. Carignan, "A comparison of acoustic and articulatory methods for analyzing vowel differences across dialects: across dialects: Data from American and Australian English," *J. Acoust. Soc. Am.*, vol. 142, pp. 363–377, 2017.

[11] J. Dang, M. Tiede, and J. Yuan, "Comparison of Vowel Structures of Japanese and English in Articulatory and Auditory Spaces," in *INTERSPEECH 2009 – 10th Annual Conference of the International Speech Communication Association*, 2009, no. 4, pp. 2815–2818.

[12] D. H. Whalen, W. Chen, M. K. Tiede, and H. Nam, "Variability of articulator positions and formants across nine English vowels," *J. Phon.*, vol. 68, pp. 1–14, 2019.

[13] S. Prom-On, P. Birkholz, and Y. Xu, "Training an articulatory synthesizer with continuous acoustic data," in *INTERSPEECH 2013 – 14th Annual Conference of the International Speech Communication Association*, 2013, pp. 349–353.

[14] S. Prom-On, P. Birkholz, and Y. Xu, "Identifying underlying articulatory targets of Thai vowels from acoustic data based on an analysis-by-synthesis approach," *Eurasip J. Audio, Speech, Music Process.*, vol. 23, 2014.

[15] A. Xu, P. Birkholz, and Y. Xu, "Coarticulation as synchronized dimension-specific sequential target approximation: An articulatory synthesis simulation," in *Proceedings of the 19th International Congress of Phonetic Sciences*, 2019, pp. 205–209.

[16] D. van Niekerk, A. Xu, B. Gerazov, P. K. Krug, P. Birkholz, and Y. Xu, "Finding Intelligible Consonant-Vowel Sounds Using High-Quality Articulatory Synthesis," in *INTERSPEECH 2020 – 21th Annual Conference of the International Speech Communication Association*, 2020, pp. 4457-4461.

[17] P. Birkholz, "Modeling Consonant-Vowel Coarticulation for Articulatory Speech Synthesis," *PLoS One*, vol. 8, no. 4, p. e60603, https://doi.org/10.1371/journal.pone.0060603, 2013.

[18] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by Simulated Annealing," *Science*, vol. 220, no.4598, pp. 671– 680, 1983.

[19] S. B. Davis and P. Mermelstein, "Comparison of Parametric Representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Trans. Acoust.*, vol. 28, no. 4, pp. 357–366, 1980.

[20] S. Prom-on and Y. Xu, "PENTATrainer2: A hypothesis- driven prosody modeling tool," in *Proceedings of Exling*, 2012.

[21] Y. Xu and H. Gao, "FormantPro as a Tool for Speech Analysis and Segmentation," *Rev. Estud. Da Ling.*, vol. 26, no. 4, pp. 1435–1454, 2018.

[22] J. L. Fleiss, B. Levin, and M. C. Paik, *Statistical Methods for Rates and Proportions*, 3rd ed. John Wiley & Sons, Inc., 2003.

[23] B. G. Evans and P. Iverson, "Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences," *J. Acoust. Soc. Am.*, vol. 115, no. 352, 2004.

[24] L. Ménard, C. Toupin, S. R. Baum, S. Drouin, J. Aubin, and M. Tiede, "Acoustic and articulatory analysis of French vowels produced by congenitally blind adults and sighted adults," *J. Acoust. Soc. Am.*, vol. 134, no. 4, pp. 2975–2987, 2013.

[25] S. R. Moisik, J. H. Esling, L. Crevier-buchman, and P. Halimi, "Putting the larynx in the vowel space: Studying larynx state across vowel quality using MRI," in *Proceedings of 19th International Congress of Phonetic Sciences (ICPhS)*, 2019.

[26] S. Wood, "The acoustical significance of tongue, lip, and larynx maneuvers in rounded palatal vowels," *J. Acoust. Soc. Am.*, vol. 80, no. 2, pp. 391–401, 1986.

[27] A. S. Mefferd, "Tongue- and Jaw-Specific Contributions to Acoustic Vowel Contrast Changes in the Diphthong /ai/ in Response to Slow, Loud, and Clear Speech," *J. Speech, Lang. Hear. Res.*, vol. 60, pp. 3144–3158, 2017.

[28] K. Otomo and C. Stoel-gammon, "The Acquisition of Unrounded Vowels in English," *J. Speech Hear. Res.*, vol. 35, pp. 604–616, 1992.

[29] B. Galantucci, C. A. Fowler, and M. T. Turvey, "The motor theory of speech perception reviewed," vol. 13, no. 3, pp. 361–377, 2006.