УДК 130.2
DOI 10.17726/phillT.2023.2.4

Нейросетевые методы классификации гласных в вокалических системах с признаком продвинутости корня языка, или ATR

Макеева Надежда Владимировна,
кандидат филологических наук,
старший научный сотрудник отдела африканских языков,
Институт языкознания РАН
Москва, Россия

umuta11@yandex.ru

Аннотация. Статья посвящена обсуждению результатов тестирования нейронной сети, классифицирующей гласные вокалической системы, включающей признак продвинутости корня языка, или [ATR] (Advanced Tongue Root), на материале языка акебу (семья ква). Акустическая природа признака [ATR] исследована недостаточно. Единственный надежный акустический коррелят [ATR] – величина первой форманты (F1) – является также акустическим коррелятом признака подъема, что создает существенные области пересечения между верхними гласными [-ATR] и средними гласными [+ATR]. Другие параметры, предлагаемые в качестве акустических коррелятов [ATR]: ширина полосы F1 (B1), разность амплитуд первых двух формант (A1-A2) и др., – различают, как правило, только часть гласных части носителей. Для обучения и тестирования нейронной сети были использованы значения четырех переменных: F1, F2, A1-A2, B1. Были протестированы четыре варианта модели, различающиеся отсутствием/наличием пятой переменной, кодирующей спикера, и количеством скрытых слоев (1 vs. 2). Модели, включающие параметр спикера, показали несколько более высокий уровень эффективности в виде доли правильных ответов, при этом показатели точности и полноты трехслойной модели оказались в целом выше, чем для модели, имеющей два скрытых слоя.

Ключевые слова: признак продвинутости корня языка; ATR; акебу; ква; нейронные сети.

Neural network methods for vowel classification in the vocalic systems with the [ATR] (Advanced Tongue Root) contrast

Makeeva Nadezhda Vladimirovna,

Candidate of Philology,
Senior Researcher, Department of African Languages,
Institute of Linguistics RAS
Moscow, Russia

umuta11@yandex.ru

Abstract. The paper aims to discuss the results of testing a neural network which classifies the vowels of the vocalic system with the [ATR] (Advanced Tongue Root) contrast based on the data of Akebu (Kwa family). The acoustic nature of the [ATR] feature is yet understudied. The only reliable acoustic correlate of [ATR] is the magnitude of the first formant (F1) which can be also modulated by tongue height, resulting in significant overlap between high [-ATR] vowels and mid [+ATR] vowels. Other acoustic metrics which had been associated with the [ATR], such as F1 bandwidth (B1), relative intensity of F1 to F2 (A1-A2), etc., are typically inconsistent across vowel types and speakers. The values of four metrics – F1, F2, A1-A2, B1 – were used for training and testing the neural network. We tested four versions of the model differing in the presence of the fifth variable encoding the speaker and the number of hidden layers. The models which included the variable encoding the speaker achieved slightly higher accuracy, whereas the precision and recall metrics of the three-layer model were generally higher than those with two hidden layers.

Keywords: Advanced Tongue Root; ATR; Akebu; Kwa; neural networks.

Введение

Признак продвинутости корня языка, или [ATR] (Advanced Tongue Root), широко распространен в нигеро-конголезских и нило-сахарских языках макросуданского пояса (Güldemann, 2008, 2010). Артикуляторная база и акустические корреляты этого признака активно изучались начиная с 1960-х годов. Изучение артикуляторной базы методами рентгенографии (Ladefoged, 1964; Painter, 1973; Lindau, 1975, 1979; Сурканова 1978), магнитно-резонансной томографии (Tiede, 1996), эндоскопической ларингоскопии (Edmondson & Esling, 2006) и ультразвуковыми методами

(Gick et al., 2006; Allen et al., 2013; Hudu, 2014; Kirkham & Nance, 2017) привело к относительно уверенному пониманию тех артикуляторных механизмов, которые обеспечивают контраст по признаку продвинутости корня языка. Особый вклад в это понимание был внесен исследованиями методом эндоскопической ларингоскопии, проводившимися с начала 2000-х годов. Эти исследования показали, что главным артикуляторным движением, работающим на создание контраста по признаку [ATR], является черпалонадгортанное сжатие, тогда как другие известные механизмы, связанные с данным контрастом, такие как движение корня языка, вертикальное смещение гортани, изменение объема глотки, являются лишь вспомогательными механизмами, работающими сообща в рамках единой синергетической системы ларингального артикулятора (Edmondson & Esling, 2006; Esling et al., 2019).

В то же время акустическая природа признака остается исследованной недостаточно. Несмотря на длительную историю исследований, был обнаружен лишь один надежный акустический коррелят признака [ATR] – величина первой форманты (F1) (Lindau, 1975; Fulop et al., 1998; Guion et al., 2004; Starwalt, 2008), являющаяся одновременно основным акустическим коррелятом признака подъема. Все попытки найти какие-либо дополнительные параметры, различающие гласные, противопоставленные по признаку [ATR], не увенчались успехом. Все предлагавшиеся параметры, такие как ширина полосы F1 (B1) (Hess, 1992; Starwalt, 2008), разность амплитуд первых двух формант (A1-A2) (Fulop et al., 1998; Guion et al., 2004; Starwalt, 2008), среднее спектральное (Starwalt, 2008; Ivanova, 2021), уровень звучности (Olejarczuk et al., 2019) и др., не показали стабильных результатов: статистически значимые различия обнаруживаются, как правило, только для части гласных части носителей.

Постановка проблемы

Обозначенное выше положение дел заставляет задаться вопросом, достаточно ли значения основного акустического коррелята признака [ATR] – F1, дополненного значениями ряда вторичных акустических параметров, таких как F2, B1 и A1-A2, для различения гласных в вокалических системах с параллельным противопоставлением по признакам подъема и [ATR], т.е. таких системах,

где имеется противопоставление по признаку [ATR] для гласных как среднего, так и верхнего подъема. Могут ли эти параметры быть использованы при описании языка для установления фонологического облика лексических единиц? Можно ли на их основе смоделировать процесс распознавания гласных в системах данного типа носителями языка? В работе будет осуществлена попытка построения нейросетевой модели, классифицирующей гласные языка акебу, относящегося к языковой семье ква (Того, 70 тыс. носителей).

Язык акебу имеет вокалическую систему, состоящую из 11 гласных. Все гласные разбиваются на два набора: [+ATR] и [-ATR]. Передние и задние гласные противопоставлены по признакам подъема и [ATR], в то время как все центральные гласные относятся к набору [-ATR], различаясь только по признаку подъема.

Таблица 1

Система гласных языка акебу

		передние	центральные	задние
верхние	+ATR	i		u
	-ATR	ɪ	ɨ	ʊ
средние	+ATR	e		o
	-ATR	ɛ	ə	ɔ
нижние	-ATR		a	

Как было упомянуто выше, основным акустическим коррелятом признака [ATR] является величина первой форманты F1. Ее значение обусловлено отсутствием или наличием черпалонадгортанного сжатия, которое при производстве гласных набора [-ATR] создает сужение в области пучности давления F1, а также способствует уменьшению длины гортанной трубки совместно с механизмом поднятия гортани (Edmondson & Esling, 2006; Edmondson, 2008). В связи с этим гласные [-ATR] имеют более высокие значения частоты F1 по сравнению с гласными [+ATR].

Однако F1 является одновременно акустическим коррелятом признака подъема. Более открытые гласные, характеризуемые меньшей степенью палатального и большей степенью фарингального сужения, имеют более высокие значения частоты F1 по срав-

нению с закрытыми гласными. Таким образом, пары гласных [i] vs. [e] и [u] vs. [o], противопоставленных по признаку подъема, как и пары гласных [i] vs. [ɪ], [u] vs. [ʊ], противопоставленных по признаку [ATR], с акустической точки зрения будут различаться величиной F1 при следующем соотношении:

F1(+ATR) < F1(-ATR) F1(i) < F1(ɪ), F1(u) < F1(ʊ);

F1(high) < F1(mid) F1(i) < F1(e), F1(u) < F1(o).

Результатом сходства акустического эффекта, достигаемого за счет противопоставления по признаку подъема с одной стороны и признака [ATR] с другой стороны, является существенное пересечение значений F1 для верхних гласных набора [-ATR] и средних гласных набора [+ATR]. Так, на формантной картине гласных языка акебу для информантов AD и AK можно наблюдать пересечение множеств значений F1 для гласных [ɪ] и [e] (рисунок 1) и гласных [ʊ] и [o] (рисунок 2) соответственно.

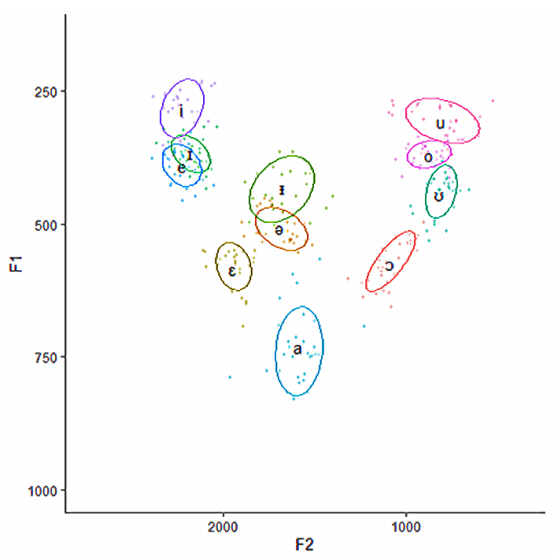

Рисунок 1. Формантная картина гласных F1-F2 (AD)

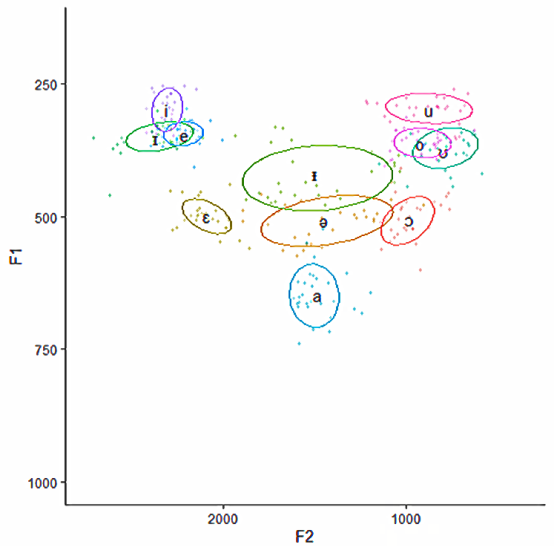

Рисунок 2. Формантная картина гласных F1-F2 (AK)

В то же время, как показало инструментально-фонетическое исследование (Makeeva & Kuznetsova, 2022), дополнительные параметры, такие как F2, B1, A1-A2, в отличие от параметра F1, не могут считаться надежными акустическими коррелятами признака [ATR] в акебу, так как не показывают стабильных статистически значимых различий для пар гласных, противопоставленных по признаку [ATR].

Методология

Материалом для исследования послужили данные языка акебу, собранные от шести информантов мужского пола в ходе полевой работы в деревне Джон в январе-феврале 2019 года: Achille Djenou (AD, 1996 г.р.), Yao Lolonyo Akossu (AK, 1967 г.р.), Koffi German Boukate (BOUK, 1958 г.р.), Kokou Mawuwodo (Honoré) Kokoroko (HO, 1990 г.р.), Kokouvi Kpoliatowou (Martin) Kodjovi (MA, 1986 г.р.), Yaovi Modeste Tchitche (YT, 1991 г.р.). Информантам было предложено трижды произнести каждое слово из списка, включавшего по 10 слов с каждым из 11 гласных языка акебу. Та-

ким образом, с учетом отбраковки около 1 процента произнесений общее число произнесений составило 1955 единиц: 10 слов * 3 повторения * 11 гласных * 6 информантов – 25.

Запись производилась при помощи профессионального цифрового аудиорегистратора Marantz PMD-660 и внешнего микрофона AKG 1000 в формате.wav при частоте дискретизации 48кГц и разрядности квантования 16 бит.

Для обучения и тестирования нейронной сети были использованы значения пяти переменных, четыре из которых представляют акустические параметры: F1, F2, A1-A2, B1. Пятая переменная, обозначающая информанта, или спикера, вводилась факультативно.

Измерения акустических параметров осуществлялись при помощи компьютерной программы анализа речи Praat (Boersma and Weenink, 2017). Измерение частоты первой форманты осуществлялось на центральном отрезке гласного, составляющем 40% от его общей длительности. Амплитуда первых двух формант и ширина полосы первой форманты измерялись на центральном отрезке гласного долготой в 0,03 сек. Перед построением модели значения акустических параметров подвергались нормализации. Значения первых двух формант (F1 и F2) были нормализованы по методу Т. Нири (Nearey, 1978, с. 138) с использованием формулы CLIH 2 (constant log interval hypothesis), позволяющей нивелировать различия между спикерами, связанные с индивидуальными особенностями артикуляторного аппарата. Значения разницы амплитуд первых двух формант (A1-A2) были нормализованы в соответствии с алгоритмом, предложенным в (Fulop et al., 1998, с. 88-91), значения ширины полосы первой форманты (B1) – при помощи метода, изложенного в (Starwalt, 2008, с. 85-90).

При помощи кода на языке Python с использованием библиотеки Keras были построены четыре варианта нейросетевой модели, различающиеся между собой отсутствием/наличием пятой переменной, кодирующей спикера, и количеством скрытых слоев (1 vs. 2):

1а. трехслойная нейронная сеть, не учитывающая спикера в качестве переменной;

1б. четырехслойная нейронная сеть, не учитывающая спикера в качестве переменной;

2а. трехслойная нейронная сеть, учитывающая спикера в качестве переменной;

2б. четырехслойная нейронная сеть, учитывающая спикера в качестве переменной.

Выбор числа нейронов в скрытых слоях опирался на геометрическое правило пирамиды. Согласно данному правилу, число нейронов в скрытом слое трехслойной нейронной сети определялось по следующей формуле:

$$k = \sqrt{n \times m},$$

где k – число нейронов в скрытом слое, n – число нейронов во входном слое, m – число нейронов в выходном слое.

В четырехслойной модели число нейронов в скрытых слоях определялось по формуле:

$$k_i = m\left(\sqrt[3]{\frac{n}{m}}\right)^{3-i}, \text{ i} = 1, 2,$$

где k_i – число нейронов в первом и втором скрытых слоях, n – число нейронов во входном слое, m – число нейронов в выходном слое.

Таблица 2

Архитектура различных типов нейросетевой модели

число нейронов / модель	входной слой (n)	первый скрытый слой (k1)	второй скрытый слой (k2)	выходной слой (m)
1а	4	7		11
1б	4	6	8	11
2а	10	10		11
2б	10	10	11	11

Результаты исследования

В таблице 3 представлены результаты эффективности четырех вариантов нейросетевой модели в виде доли правильных ответов.

Таблица 3

Доля верных ответов при различных типах нейросетевой модели

1а	1б	2а	2б
0.77	0.76	0.80	0.80

Как видно из таблицы, параметр количества скрытых слоев не влияет на эффективность модели. Показатели эффективности моделей 1б и 2б, имеющих по два скрытых слоя, отличаются от показателей эффективности моделей 1а и 2а соответственно, имеющих по одному скрытому слою, незначительно или не отличаются вовсе. Напротив, введение в качестве независимой переменной параметра спикера немного повышает эффективность моделей 2а и 2б по сравнению с моделями 1а и 1б соответственно.

Исследование метрик точности и полноты для моделей 2а и 2б показывает, какие гласные вызывают наибольшие затруднения при классификации.

Таблица 4

Метрики точности и полноты для моделей 2а и 2б

метрика / гласный	2а		2б	
	точность	полнота	точность	полнота
a	0.95	0.92	0.95	0.92
i	0.74	0.86	0.75	0.89
ɪ	0.68	0.40	0.61	0.40
e	0.61	0.83	0.61	0.75
ɛ	0.94	0.97	0.91	0.97
ɨ	0.80	0.67	0.72	0.70
ə	0.71	0.76	0.74	0.74
u	1.00	0.89	0.97	0.87
ʊ	0.89	0.76	0.91	0.91
o	0.69	0.93	0.76	0.90
ɔ	0.92	0.92	0.92	0.89

Как видно из таблицы 4, обе модели наименее точно (точность < 0.70) определяют классы ɪ и e, а наименее полно – класс ɪ. Напротив, наиболее точно (> 0.90) обе модели позволяют определить классы a, ɛ, u, ɔ, а наиболее полно – классы a и ɛ. В целом при незначительных различиях значения данных двух метрик для первой – трехслойной – модели немного выше, нежели для второй, включающей два скрытых слоя.

Выводы и дискуссия

Исследование показало, что успешность классификации гласных при помощи нейросетевой модели не зависит от архитектуры нейронной сети, т.е. от количества скрытых слоев и их нейронов. Несмотря на предварительную нормализацию значений первых двух формант, помогающую нивелировать различия между спикерами, связанные с индивидуальными особенностями строения артикуляторного аппарата, эффективность модели, учитывающей в качестве переменной спикера, несколько выше эффективности модели, учитывающей в качестве переменных лишь акустические параметры.

Несмотря на довольно высокий процент правильных ответов алгоритма (около 80%), он оказывается недостаточным для распознавания гласных, в связи с чем установление фонологического облика лексических единиц при помощи нейронной сети должно быть дополнено морфонологическим анализом, включающим в частности анализ поведения гласных в качестве триггеров вокалической гармонии по признаку [ATR].

Вопрос распознавания и классификации гласных носителями языка остается на данный момент открытым и требует дальнейшего детального изучения при помощи фонетических экспериментов с одной стороны и поиска надежных акустических коррелятов признака [ATR] – с другой.

Литература

1. *Сурканова И. М.* (1978). О некоторых артикуляторно-акустических характеристиках вокализма языка ибо. Проблемы фонетики, морфологии и синтаксиса африканских языков. М.: Издательство Московского университета. 166-204.
2. *Allen B., Pulleyblank D., Ajíbóyè O.* (2013). Articulatory mapping of Yoruba vowels: An ultrasound study. Phonology, 30. 183-210.

3. *Edmondson J.A. & Esling J. H.* (2006). The valves of the throat and their functioning in tone, vocal register and stress: Laryngoscopic case studies. Phonology, 23. 157-191.
4. *Edmondson J. A.* (2008). Correspondences between articulation and acoustics for the feature [ATR]: the case of two Tibeto-Burman languages and two African languages. Ms.
5. *Esling J. H., Moisik S. R., Benner A., Crevier-Buchman L.* (2019). Voice quality: The laryngeal articulator model. Cambridge: Cambridge Univ. Press.
6. *Fulop S. A., Kari E. & Ladefoged P.* (1998). An acoustic study of the tongue root contrasts in Degema vowels. Phonetica, 1998, 55. 80-98.
7. *Gick B., Pulleyblank D., Campbell F., Mutaka N.* (2006). Low vowels and transparency in Kinande vowel harmony. Phonology, 23. 1-20.
8. *Guion et al.* 2004 – Guion S. G., Post M. W., Payne D. L. Phonetic correlates of tongue root vowel contrasts in Maa. Journal of Phonetics, 2004, 32: 517-542.
9. *Güldemann T.* (2008). The Macro-Sudan Belt: towards identifying a linguistic area in northern sub-Saharan Africa. A linguistic geography of Africa. Heine B. & Nurse D. (eds.). Cambridge: Cambridge University Press. 151-185.
10. *Güldemann T.* (2010). Sprachraum and geography: linguistic macro-areas in Africa. Handbooks of Linguistics and Communication Science, 30, Language and Space: an International Handbook of Linguistic Variation, 2: language mapping. Lameli A., Kehrein R., Rabanus S. (eds.). Berlin: Mouton de Gruyter, 561-585, Maps 2901-2914.
11. *Hess S.* (1992). Assimilatory effects in a vowel harmony system: An acoustic analysis of advanced tongue root in Akan. Journal of Phonetics, 20. 475-492.
12. *Hudu F.* (2014). [ATR] feature involves a distinct tongue root articulation: Evidence from ultrasound imaging. Lingua, 143. 36-51.
13. *Kirkham S., Nance C.* (2017). An acoustic-articulatory study of bilingual vowel production: Advanced tongue root vowels in Twi and tense/lax vowels in Ghanaian English. Journal of Phonetics, 62. 65-81.
14. *Ladefoged P.* (1964). A Phonetic Study of West African Languages: An Auditory-Instrumental Survey. (West African Language Monographs, I.) Cambridge: Cambridge University Press.
15. *Lindau M.* (1975). [Features] for vowels. UCLA Working Papers in Phonetics, 1975, 30.
16. *Lindau M.* (1979). The feature expanded. Journal of Phonetics, 7. 163-176.
17. *Makeeva, N. & Kuznetsova, N.* (2022). Phonological and acoustic properties of [ATR] in the vowel system of Akebu (Kwa). Phonology, Vol. 39, no. 4. 671-699.
18. *Nearey, T.* (1978). Phonetic Feature Systems for Vowels. Indiana University Linguistics Club, Bloomington.
19. *Olejarczuk P., Otero M. A., Baese-Berk M. M.* (2019). Acoustic correlates

of anticipatory and progressive [ATR] harmony processes in Ethiopian Komo. Journal of Phonetics, 2019, 74. 18-41.
20. *Painter C.* (1973). Cineradigraphic data on the feature ‘covered’ in Twi vowel harmony. Phonetica, 1973, 28. 97-120.
21. *Starwalt C. G. A.* (2008). The acoustic correlates of ATR harmony in seven- and nine- vowel african languages: a phonetic inquiry into phonological structure. Ph D. dissesrtation. The University of Texas at Arlington.
22. *Tiede M. K.* (1996). An MRI-based study of pharyngeal volume contrasts in Akan and English. Journal of Phonetics, 24. 399-421.