УДК 81’342
DOI: 10.26907/2782-4756-2023-71-1-10-13

# ПРИМЕНЕНИЕ ЭВКЛИДОВОЙ МЕТРИКИ ДЛЯ ИЗМЕРЕНИЯ МЕЖФОНЕМНЫХ РАССТОЯНИЙ

**© Анатолий Зельдин**

# THE EUCLIDEAN METRICS APPLIED TO THE INTERPHONEMIC DISTANCE MEASUREMENTS

**Anatoly Zeldin**

The similarity or dissimilarity of the spoken words is generally rendered by intuition, depending on the personal orientation or the personal traits of a listener/speaker. The existing methods of phonetic encoding of words suffer from a number of shortcomings, the main one being the impossibility of weighing spoken words in quantitative terms. Moreover, the existing methods may be related to a certain language or language family. The algorithm advanced in the present paper compares the characteristics of different phonemes that make up a word. The paper treats phonemic frequency and sonority as elements common both for consonants and vowels, backness and openness, as features pertaining to vowels, and the place of articulation pertaining to consonants only. The algorithm in question permits to compare in quantitative terms the words of different length, whether formed by open or closed syllables. The inter-phonemic distances are calculated by employing Euclidean metrics. The paper suggests fields of application of the method treated in the paper: this scheme can be applied in the fields of comparative linguistics, in medicine, when the hearing disorders are scrutinized, as well as in the brain cortex mapping.

*Keywords:* phoneme, consonant, vowel, frequency, sonority

Вопрос «похожести» / «непохожести» двух или более слов одного или разных языков оценивается, как правило, интуитивно, в зависимости от индивидуальных особенностей аудиторного восприятия. Существующие методы фонетического описания слов обладают рядом недостатков, основным из них является невозможность количественной оценки и сравнения различных слов. Кроме того, предложенные алгоритмы могут быть ориентированы на конкретный язык или языковую семью. Предлагаемый метод основан на количественном сравнении отдельных параметров фонем, из которых состоит слово. В качестве параметров, общих для гласных и согласных фонем, предлагается рассмотреть высоту и сонорность, характеризующих только гласные – подъем и ряд, характеризующих только согласные – место образования. Межфонемные расстояния вычисляются по Эвклидовой метрике. Рассматриваемый алгоритм позволяет сравнивать в количественных условных единицах слова различной длины, которые состоят из различных видов слогов – как закрытых, так и открытых. Предложены области применения предлагаемого метода: эта схема может применяться в области сравнительного языкознания, в медицине при изучении нарушений слуха, а также при картировании коры головного мозга.

*Ключевые слова*: фонема, согласные, гласные, частота, сонорность

*Для цитирования*: Зельдин А. Применение эвклидовой метрики для измерения межфонемных расстояний // Филология и культура. Philology and Culture. 2023. №1 (71). С. 10–13. DOI: 10.26907/2782-4756-2023-71-1-10-13

“Why, if a fish came to me, and told me he was going on a journey,
I should say “With what porpoise”?
“Don’t you mean “purpose”? – said Alice.
“I mean what I say”, the Mock Turtle replied in an offended tone”.
“Alice in the Wonderland”

As the father of the modern linguistics put it, “the first linguists, who knew nothing about the physiology of articulated sounds, were constantly falling into a trap, to me, it means a first step in the direction of truth, for the study of sounds themselves furnishes the desired prop” [2, p. 32]. As a rule, the mutual understanding or misunderstanding stems from the shared phonetic structure of a word (or a lexeme), and, to a lesser extent, from the shared semantics common for a speaker and a listener.

The word (or a lexeme) consists of phonemes. Or, citing de Saussure again “language…is a system based on the mental opposition of auditory impressions, just as a tapestry is a work of art produced by the visual oppositions of threads of different colors…” [Ibid.,p. 39]. The phonemes are classified as the consonants and vowels, we do not consider for the moment the clicks found almost exclusively in the Khoisan languages. The consonants may be pulmonic and non-pulmonic; the phonetic systems that include the non-pulmonic consonants are few and far between, and so far we exclude them from the discussion.

The main feature that differentiates the human speech from other species’ communication systems is the use of words or lexemes which consist of phonemes. A phoneme is the smallest meaningful ’sound unit’ in language (*cat* vs. *bat*, *cat* vs. *cut*, *cat* vs. *can*). The IPA defines the common consonant phonemes as shown in Fig. 1. The arrangement of rows in the table is clear – from the most outer sound (bilabials /m/, /b/ etc. moving deeper into the mouth cavity toward the ‘deepest’ ones – the glottal consonants, like /ʔ/. On the contrary, the up-down order of the consonant rows is purely arbitrary – nasals traditionally occupy the upper row, stops – the second upper row and so on.

The vowels are more strictly organized – each sound occupies the position in the vowel chart (Fig. 2) according to the tongue height at the moment of vocalization and backness (the tongue position relative to the back of mouth). A vowel position could be easily pinpointed on a chart; to map a consonant is a much harder task – only one axis coordinate is provided.

The problem is expressed in the difficulties one can face while trying to quantitatively define a phoneme and, consequently, a whole word. We suppose that the solution may bear on the distinguishing of phonetic and even phonemic proximity of different phonemes, syllables, words and expressions, and could be applied in various fields (aka comparative linguistics, neurolinguistics, auditory impairments treatment).

When a need to determine an inter-phonemic distance arises, first and foremost the Levenstein distance comes to mind [5]. Among its shortcomings are the artificialness and the over-relating to orthography. Beside this, only the letters’ order is of importance, due to this the differences in the phonetic vicinity or remoteness are wiped away. Hence, the lexemes *but* – *cut* – *hut* – *gut* – *nut* – *tut* (*tit* for *tut*) occupy the same position in terms of mutual distance when the Levenstein method is applied, although the phonemic perception in each case is different, of course. To our mind, *but* is closer to *gut* than to *cut*, and *nut* is still more far away. One should take into consideration that the relative closeness of the sounds (or phonemes) and words is a rather subjective perception and varies from one person to another.

The popular algorithm SoundEx (based on the Germanic languages and applied predominantly to the Germanic words) [4] ignores the individual sound features, as does the Levenstein distance. The same may be said of the Metaphone phonetic algorithm [8].

“Speech can be produced rapidly because the phonemes are processed in parallel. They are taken apart into their constituent featured…” [6, p. 454]. The algorithm advanced in the present paper is based on the acoustic properties of a phoneme, these properties are expressed quantitatively and taken in an array. As mentioned above, the IFA consonant table while providing the exhaustive phonetic account of the world languages, nevertheless has a significant shortcoming – an arbitrary arrangement of rows that display the manner of vocalization. Thus, the row of nasal consonants could as well be changed with the row of plosives, etc.

We can use the index of the ‘consonant deepness’ that shows the place of articulation, i.e. how deep a consonant in the vocal cavity is produced. Index 1 stands for the bilabial sounds, 2 – for the labio-velar sounds… 13 – for the glottal sounds. But for the sake of acoustic closeness evaluation of phonemes pertaining to different classes, we need at least one more quantitative parameter. As such, so called COG can be used. “The center of gravity of a spectrum (COG) is in a sense, the “mean” frequency” [10, p.1530]. Also, a phoneme sonority may be used, although determining this feature is not a simple task, albeit it is easily perceived intuitively. For example, the sonority was described as “the loudness relative to that of other sound with the same length, stress and pitch” (Ladefoged) [cit.: 3, p. 20].

To visualize the vowels and the consonants in a unitary frame of reference, we chose the features of the sonority and the frequency (the deepness is unfit for vowels, the openness and backness are used instead). The quantitative values were cited from the following sources - $S_i$ (the sonority of sound i) [3], $F_i$ (the frequency or COG of sound i) [10]. The x-axis marks sonority, the y-axis marks the mean frequency measured in Hz. So, /m/ (17; 18), /n/ (15; 19), /p/ (7; 39), /f/ (7;4), /a/ (24; 38). The distance between syllables or the whole words will be calculated according the standard Euclidian metrics:

$$D\ (A_1, A_2) = \sqrt{\sum_{i=1}^{n}(x_i^{(1)} - x_i^{(2)})^2}$$

To apply the metrics in a special case, one can transform the equation in the following way:

$$D(A_1, A_2) = \sqrt{\mu_1 (x_1 - x_2)^2 + \mu_2 (x_1 - x_2)^2 + \mu_3 (x^1 - x_2)^2},$$

where $\mu_i$ - the specific weight of every parameter; it describes the subjective or personal experience of every phoneme and could be determined by experimental approach.

However, we are facing the problem of dimensionality, i.e. the units of measurement, which we choose to describe the properties of a certain phoneme. To solve the problem of alignment acoustic and metric units, we should normalize the numbers or transform them into unidimensional units:

$$\bar{x} = \frac{1}{n} * (x_1 + x_2 + \ldots + x_n) = \frac{1}{n} * \sum_{i=1}^{n} x_i ,$$

where $\bar{x}$ is the average of a given parameter, n – the total number of phonemes in question.

If a normal distribution takes place, the equation will look like:

$$\bar{x}_i = (x_i - \bar{x}) / \sqrt{\sum_{i=1}^{n} (x_1 - \bar{x})^2}$$

The divisor is the standard deviation of a considered value. As long as we are not safe that the distribution in question is a normal one, it is safer to apply a more general case of transformation:

$$\tilde{x}_i = (x_i - x_{min}) / (x_{max} - x_{min}) \in [0,1];$$

The sonority values are: $S_{max} = 29$, $S_{min} = 0$ [Dineen & Miller 1995, 20]; the mean frequency values: $F_{max} = 400$ Hz, $F_{min} = 18$ Hz [van Son & Pols 1996, 1531]; the place of articulation (or “deepness”) relative to the most front consonant: $D_{max} = 13$, $D_{min} = 1$.

Therefore: $\bar{S}_m = (S_m - S_{min}) / (S_{max} - S_{min}) = 0.59$; $\bar{S}_n = 0{,}52$; $\bar{S}_p = 0{,}244$; $\bar{S}_f = 0{,}24$; $\bar{S}_a = 0{,}83$;

The same about COG or the mean frequency: $\bar{F}_m = (F_m - F_{min}) / (F_{max} - F_{min}) = 0.02$; $\bar{F}_n = 0.023$; $\bar{F}_p = 0.077$; $\bar{F}_f = 1$; $\bar{F}_a = 0.077$;

The relative deepness of a phoneme or the place of articulation: $\bar{D}_m = (D_m - D_{min}) / (D_{max} - D_{min}) = 0.077$; $\bar{D}_n = 0.307$; $\bar{D}_p = 0.077$; $\bar{D}_f = 0.154$;

As for the vowels we can use the axis of backness (Fig. 2) – values 0, 1, 2, 3 or 4, and openness – values from 0 to 8. So, $B_a = 1$, $\bar{B}_a = 0.25$; $O_a = 1$, $\bar{O}_a = 0.125$;

Ultimately, the phonemic matrix of *mama* word will look like: $[\bar{S}_m|\bar{F}_m|\bar{D}_m;\ \bar{S}_a|\bar{F}_a|\bar{B}_a|\bar{O}_a;\ \bar{S}_m|\bar{F}_m|\bar{D}_m;\ \bar{S}_a|\bar{F}_a|\bar{B}_a|\bar{O}_a]$ = [0.59|0.02|0.077; 0.83|0.077|0.25|0.125; 0.59|0.02|0.077; 0.83|0.077|??]. The matrix of *nana*: [0.52|0.023|0.307; 0.83|0.077|0.25|0.125; 0.52|0.023|0.307; 0.83|0.077|0.25|0.125], of *papa*: [0.24|0.077|0.077; 0.83|0.077|0.25|0.125; 0.24|0.077|0.077; 0.83|0.077|0.25|0.125], of *fafa*: [0.24|0.95|0.154; 0.83|0.077|0.25|0.125; 0.24|0.95|0.154; 0.83|0.077|0.25|0.125].

The inter-word distances finally should be calculated in the following way:

$$d(\textit{mama}, \textit{nana}) = \sqrt{(\bar{S}_m - \bar{S}_n)^2 + (\bar{F}_m - \bar{F}_n)^2 + (\bar{D}_m - \bar{D}_n)^2 + (\bar{S}_a - \bar{S}_a)^2 + (\bar{F}_a - \bar{F}_a)^2 + (\bar{B}_a - \bar{B}_a)^2 + (\bar{O}_a - \bar{O}_a)^2 + (\bar{S}_m - \bar{S}_n)^2 + (\bar{F}_m - \bar{F}_n)^2 + (\bar{D}_m - \bar{D}_n)^2 + (\bar{S}_a - \bar{S}_a)^2 + (\bar{F}_a - \bar{F}_a)^2 + (\bar{B}_a - \bar{B}_a)^2 + (\bar{O}_a - \bar{O}_a)^2} = \sqrt{(0.59 - 0.24)^2 + (0.02 - 0.077)^2 + (0.077 - 0.077)^2 + (0.83 - 0.83)^2 + (0.077 - 0.077)^2 + (0.25 - 0.25)^2 + (0.125 - 0.125)^2 + (0.59 - 0.24)^2 + (0.02 - 0.077)^2 + (0.077 - 0.077)^2 + (0.83 - 0.83)^2 + (0.077 - 0.077)^2 + (0.25 - 0.25)^2 + (0.125 - 0.125)^2} = \sqrt{0.11578} = 0.342$$

d (*mama*, *papa*) = 0.501

d (*mama*, *fafa*) = 1.409

As one might expect, the *mama* word phonetically (or phonologically) is closer to *nana* than to *papa*, and the more so, to *fafa*. Along with this, the phonetic closeness and remoteness can be easily compared. Hence, an **abcd* ($A_1$) word would be twice closer to **efgh* ($A_2$) than to **ijkl* ($A_3$). Of course, the vowels are not to be neglected. Thus, we may compare *mama* with *mimi*. $\bar{S}_i = 0.758$, $\bar{F}_i = 0.077$; $O_a = 1$, $\bar{O}_a = 0.125$, $B_a = 1$, $\bar{B}_a = 0.25$; $O_i = 8$, $\bar{O}_i = 1$, $B_i = 4$, $\bar{B}_i = 1$. Hence, d(*mama*, *mimi*) = 1.63. So, we can infer that these words seem rather dissimilar by the auditory perception.

In conclusion, an obvious question must be asked: how C (a consonant) and V (a vowel) are to be compared? In other words, when there is no evident syllables’ alignment like CVC-CVC or VCV-VCV, how could the inter-word distance between lexemes like CV and VC be calculated? We deem the comparing with zero would be the best way to tackle the problem. Let’s consider two words $A_1$- CV (or $C_1V_1$), $A_2$ – VC (or $V_2C_2$), like *be* and *of*, to be compared. The distance required should be calculated in the following way: $D(A_1, A_2) = \sqrt{(\bar{S}_{c1} - \bar{S}_{v2})^2 + (\bar{F}_{c1} - \bar{F}_{v2})^2 + (\bar{D}_{c1} - 0)^2 + (0 - \bar{O}_{v2})^2 + (0 - \bar{B}_{v2})^2 + (\bar{S}_{v1} - \bar{S}_{c2})^2 + (\bar{F}_{v1} - \bar{F}_{c2})^2 + (0 - \bar{D}_{c2})^2 + (\bar{O}_{v1} - 0)^2 + (\bar{B}_{v1} - 0)^2}$;

The same may be applied to the words of different length. For example, if $V_1C_1V_2$ and $C_2V_3$ (like *era* and *to* words) are considered, the equation will look like this:

$$d(A_1, A_2) = \sqrt{(\bar{S}_{v1} - \bar{S}_{c2})^2 + (\bar{F}_{v1} - \bar{F}_{c2})^2 + (0 - \bar{D}_{c2})^2 + (\bar{O}_{v1} - 0)^2 + (\bar{B}_{v1} - 0)^2 + (\bar{S}_{c1} - \bar{S}_{v3})^2 + (\bar{F}_{c1} - \bar{F}_{v3}) + (\bar{D}_{c1} - 0) + (0 - \bar{O}_{v3})^2 + (0 - \bar{B}_{v3})^2 + (\bar{S}_{v2} - 0)^2 + (\bar{F}_{v2} - 0)^2 - (\bar{O}_{v2} - 0)^2 + (\bar{B}_{v2} - 0)^2}$$

**Conclusions**

The outlined method is just a preliminary proposition. We do not consider so far such features as the specific weight of different sound properties (labeled $\mu_i$) at their phonetic perception, the impact of a phoneme surrounding [6] and the stressed/unstressed syllable, the vowel longevity, tone, etc. The laboratory experiments for the refinement of the proposed method are needed. Besides this, the method in its present form seems to be a little cumbersome, demanding multiple and tedious calculations, but when an appropriate software is developed, the task of inter-word distance calculations will become much easier and ready-to-hand like the determining of the semantic distance according the WordNet algorithm [7].

The advanced scheme can be applied in the fields of comparative linguistics, especially in the case when a phoneme shift like *satum-kentum* is considered in the historic perspective, in medicine, when the hearing and hearing disorders are scrutinized, as well as in the brain cortex mapping, following the phonotopic principle, according to which, every phoneme pins a certain point or a patch of cortex [9].

References


1. Crystal, D. (1987). *The Linguistic Use of Sound*. The Cambridge Encyclopedia of Language. Cambridge University Press. (In English)
2. De Saussure, F. (1959). *Course in General Linguistics*. New York, The Philosophical Library. (In English)
3. Dineen, N., Miller, D. (1995). *The Derivation of a Sonority Hierarchy from the Syllable Contact Law (SCL) and the Productivity of the SCL in American English*. George Mason University Working Papers in Linguistics. Vol. 5, pp. 19–47. (In English)
4. Knuth, D. (1973). *The Art of Computer Programming*. V. 3: Sorting and Searching. Addison-Wesley. (In English)
5. Levenshtein, V. (1966). *Binary Codes Capable of Correcting Deletions, Insertions, and Reversals*. Soviet Physics Doklady. No. 10 (8), pp. 707–710. (In English)
6. Lieberman, A. M., Cooper, F. S., Shankweiler, D. P., Studdert-Kennedy, M. (1967). *Perception of the Speech Code.* Psychological Review. No. 74(6), pp. 431-461. (In English)
7. Pedersen, T., Patwardhan, S., Michelizzi, J. (2004). *WordNet:Similarity – Measuring the Relatedness of Concepts*. Proceedings of the XIX Nat. Conf. on Artificial Intelligence (AAAI-04). July 25-29, pp.1024-1025 San Jose, CA (Intelligent Systems Demonstration). (In English)
8. Philips, L. (1990). *Hanging on the Metaphone.* Computer Language Magazine. No. 7 (12), pp. 39-44. (In English)
9. Pulvermüller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O. (2006). *Motor cortex map articulatory features of speech sounds.* PNAS. No. 103(20), pp. 7865-7870. (In English)
10. Van Son, R .J. J. H., Pols, L. C. W. (1996). *An Acoustic Profile of Consonant Reduction*. Proceeding of Fourth International Conference on Spoken Language Processing ICSLP'96. Pp. 1529-1532. (In English)
11. Vinogradov, V. (1990). *Fonema* [*The Phoneme*]. Lingvisticheskii Entsiklopedicheskii Slovar'. Yartseva V. (ed.-in-chief), Sovetskaya Entsiklopediya, Pp. 552-554. Moscow. (In Russian)


The article was submitted on 10.09.2022
Поступила в редакцию 10.09.2022

**Зельдин Анатолий**,
соискатель,
Управление государственной службы,
76300, Реховот, Израиль.
anatolyz@moia.gov.il

**Zeldin Anatoly**,
Ph.D. applicant,
Civil Service Commission,
Rehovot, 76300, Israel.
anatolyz@moia.gov.il