

Аналіз звуків мови в програмі Praat

Паренюк А. В., ORCID [0000-0002-8882-5976](https://orcid.org/0000-0002-8882-5976)

Кафедра акустичних та мультимедійних електронних систем, ames.kpi.ua
Національний технічний університет України
«Київський політехнічний інститут ім. Ігоря Сікорського», ROR [00syn5v21](https://ror.org/00syn5v21)
Київ, Україна

Анотація—У статті визначена можливість використання програми Praat для комплексного і повного аналізу звуків мови. Зазначено спектр функцій програмного комплексу та підкреслено перспективу використання програми для порівняння звуків. Було проаналізовано спектрограми, проведено формантний аналіз, аналіз інтенсивності, оцінено низку інших параметрів, таких як Jitter, Shimmer, розриви голосу. На основі отриманих результатів відбулося порівняння обраних звуків. У підсумку був зроблений висновок, щодо доцільності і ефективності використання програмного комплексу Praat при аналізі звуків мови.

Ключові слова — спектрограма; форманта; висота тону звуку; інтенсивність; Jitter (тремтіння); Shimmer (мерехтіння); розриви голосу; звук; програмний комплекс Praat; частота.

I. ВСТУП

Мова – це сукупність довільно відтворюваних загальноприйнятих у межах даного суспільства звукових знаків для об'єктивно існуючих явищ і понять, а також загальноприйнятих правил їх комбінування у процесі вираження думок[1]. Мовлення — це здатність говорити, активність розмови або частина розмовної мови[2]. Мова – це важливий засіб комунікації, саме тому аналіз звуків мови є важливим напрямом дослідження, оскільки результати цих досліджень можуть використовуватися в різноманітних сферах людської діяльності, наприклад, лінгвістиці або фоніатрії (галузь оториноларингології, яка займається діагностикою, лікуванням та профілактикою з ахворювань голосоутворюючої системи [3]).

Саме програмний комплекс Praat, перше знайомство з яким відбулося в рамках профільної дисципліни і який буде розглянуто нижче, може бути вказаний, як ефективний засіб, що задовольняє поставлену задачу і був використаний в даній роботі. Мета роботи — обґрунтування доцільності використання програмного комплексу Praat для аналізу мови шляхом порівняння обраних звуків на основі формантного аналізу, аналізу інтенсивності, оцінки висоти тону звуку, аналізу акустичних характеристик jitter та shimmer, а також аналіз самих звуків мови і їх порівняння.

II. МЕТОДИ ДОСЛІДЖЕНЬ

Praat — це програма для аналізу та синтезу мовлення, написана Полом Боерсма та Девідом Венінком на кафедрі фонетики Амстердамського університету.

Praat — це дуже гнучкий інструмент для аналізу мовлення. Він пропонує широкий спектр стандартних та нестандартних процедур.

Praat призначений для тих, хто хоче аналізувати, синтезувати і маніпулювати мовленням. Крім того, за допомогою цього інструменту можна створювати

високоякісні рисунки, які можливо експортувати і вставляти в наукові статті або особисті дослідження. Слід зазначити, що існують також інші програми, що використовуються для вирішення аналогічних завдань, такі як Audacity, Phonology Assistant IPA Help та Speech Analyzer, проте дана програма є абсолютно безкоштовною, з простим графічним інтерфейсом та можливістю в будь-який момент скористатися довідкою, яка викликається в вікні програми.

Далі розглянемо можливості програмного комплексу.

Аналіз мови:

- спектрограми;
- формантний аналіз;
- аналіз інтенсивності;
- Jitter, Shimmer, розриви голосу;
- кохлеограма;
- картина збудження.

Синтез мови:

- від тангажу, формант і інтенсивності;
- артикуляційний синтез;
- акустичний синтез Клатт.

Маркування та сегментація:

- використання фонетичного алфавіту;
- використання звукових файлів об'ємом до 2 гігабайт (3 години).

Мовні маніпуляції:

- зміна висоти тону і тривалості контурів;
- фільтрація.



Статистика:

- багатовимірне обчислення;
- аналіз головних компонентів;
- дискримінантний аналіз [4].

Більшість операцій, які проводять із Praat, починаються зі звуку. Існує принаймні три способи отримання звуку в програмі:

- запис звуку
- зчитування звуку з диска
- створення звуку з формули

В даній роботі використовувався перший спосіб і розглядалися голосні звуки «а» та «е», а також приголосні «н» та «п». Приголосні звуки обиралися так, щоб була можливість попрацювати, як з дзвінким приголосним, так і з глухим, проте загального підходу або якогось правила до обрання звуків не було, оскільки метою було показати саме ефективність програми в умовах вільного вибору користувачем вхідних даних.

III. РЕЗУЛЬТАТИ

Розглянемо спершу голосні звуки попередньо видаляючи неінформативні частини запису.

Наведемо також зображення спектрограми. Спектрограма показує перетворення Фур'є сигналу, що змінюється в часі. Величина частотних компонентів зазвичай або представлена як змінюються кольори (по заданій шкалі кольорів), або як різні відтінки чорного для графіка в градаціях сірого [5].

Аналізуючи рис. 2 можна зробити висновок, що в більш затемнених зонах розташовуються формантні смуги. Наразі розглянемо ці смуги, що також є можливим в даному програмному комплексі.

Як видно з рис. 3 ми маємо 4 форманти, оскільки було використано стандартне налаштування, що в загальному є достатнім для роботи, проте ці налаштування в будь-який момент можуть бути змінені. Слід також зауважити, що максимальний поріг діапазону пошуку формант за замовчуванням встановлений для жінки-мовця [6]. Також можна отримати повну інформацію по наявних формантах:

- F1_Hz (mean) = 887.5486257770023 Hz
- F2_Hz (mean) = 1248.5970943045884 Hz
- F3_Hz (mean) = 3823.684300648018 Hz
- F4_Hz (mean) = 4182.399262534927 Hz

Різниця між 2 та 3 формантою становить більше 2 кГц.

Що в певній мірі відповідає теоретичним даним [7], проте з певними відмінностями, що є цілком передбачуваним, оскільки для більш точних результатів слід провести значну кількість дослідів.

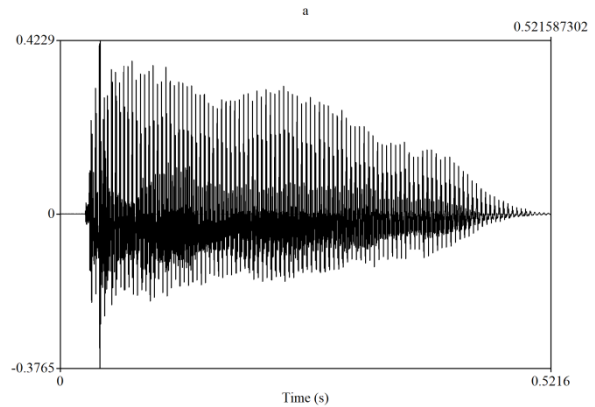


Рис. 1 Форма заданого сигналу (звук «а»)

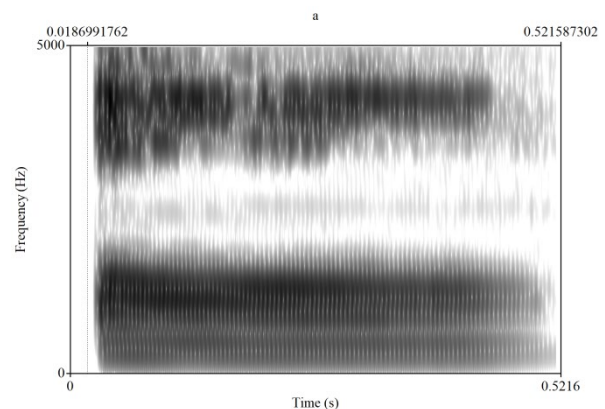


Рис. 2 Спектрограма заданого сигналу (звук «а»)

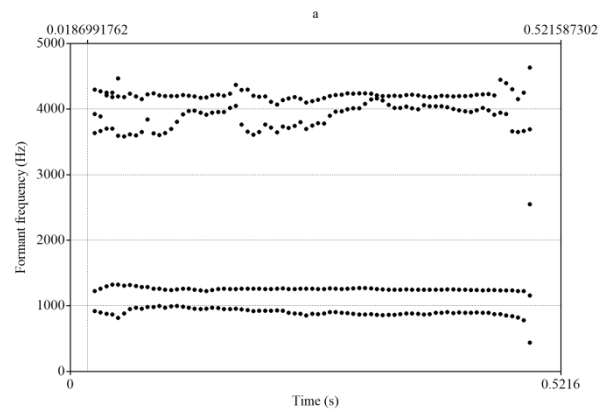


Рис. 3 Зображення формант заданого сигналу (звук «а»)

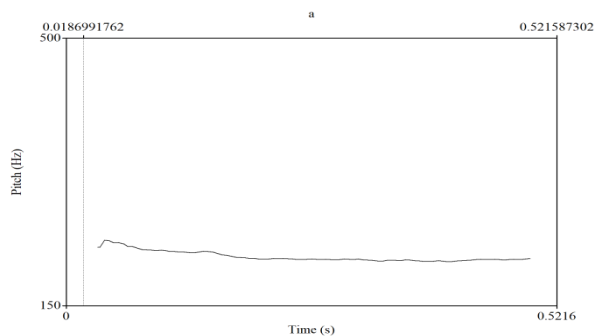


Рис. 4 Зображення висоти тону заданого сигналу (звук «а»)

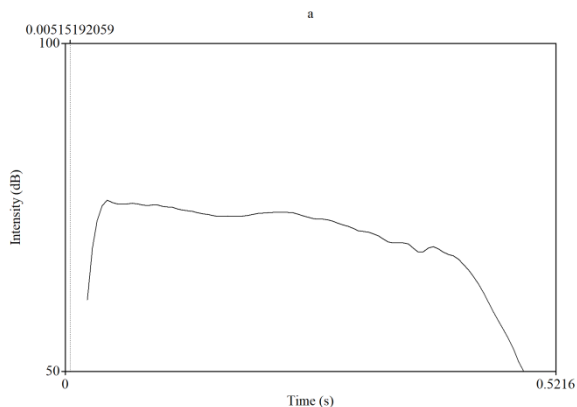


Рис. 5 Зображення інтенсивності заданого сигналу (звук «а»)

Далі оцінимо висоту тону звуку, що визначається об'єктивною характеристикою — частотою (чим більша частота тим вищий тон звуку при його слуховому сприйнятті)[8].

Було також отримано мінімальну і максимальну висоту тону, а також середнє значення:

- minimum pitch: 207.51183356609093 Hz
- maximum pitch: 236.67825551165137 Hz
- mean pitch: 214.33337813636768 Hz

Розглянемо також інтенсивність звуку.

Визначимо мінімальну і максимальну інтенсивність, а також її середнє значення:

- minimum intensity: 46.29862369716834 dB
- maximum intensity: 76.13476148710619 dB
- mean-energy intensity: 72.5417405248146 dB

Також було отримано загальний звіт щодо даного сигналу (Jitter, Shimmer і розриви голосу):

Voicing:

- Fraction of locally unvoiced frames: 7.071% (7/99)
- Number of voice breaks: 0
- Degree of voice breaks: 0 (0 seconds / 0 seconds)

Jitter:

- Jitter (local): 0.430%
- Jitter (local, absolute): 20.045E-6 seconds

Shimmer:

- Shimmer (local): 4.431%
- Shimmer (local, dB): 0.497 dB

Jitter та shimmer — це акустичні характеристики голосових сигналів і в основному використовуються для виявлення патологій голосового апарату[9].

Наступним розглянемо звук «е».

Очевидно, що звуки різняться за формою в чому ми можемо наочно переконались. Далі наведемо спектрограму сигналу.

Як видно з рис. 7, ми можемо візуально виділити 4 більш затемнені зони. Далі розглянемо форманти даного звуку.

Очевидно, що отримане зображення повністю відрізняється від зображення для звуку «а». Наведемо повну інформацію по наявних формантах:

- F1_Hz (mean) = 622.8977924451594 Hz
- F2_Hz (mean) = 1761.0960438226905 Hz

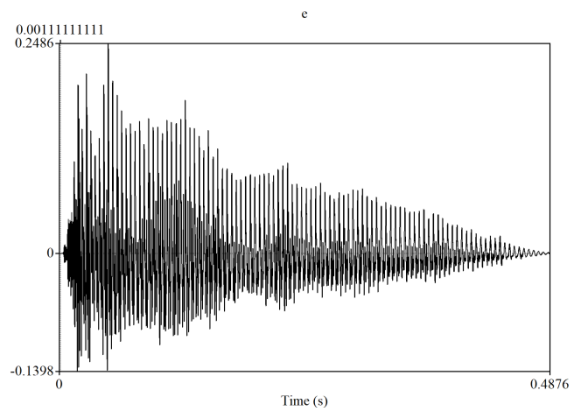


Рис. 6 Форма заданого сигналу (звук «е»)

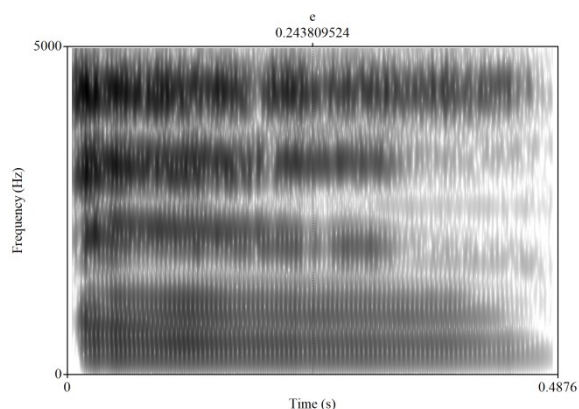


Рис. 7 Спектрограма заданого сигналу (звук «е»)

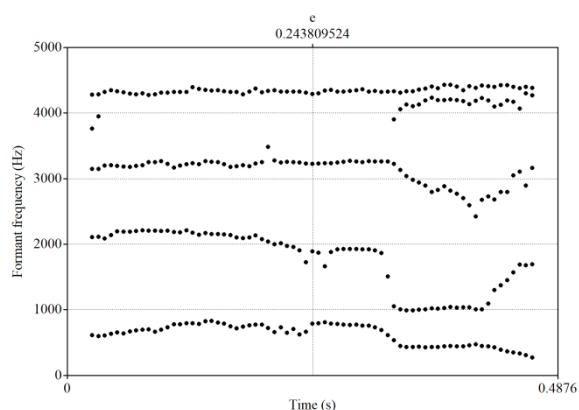


Рис. 8 Зображення формант заданого сигналу (звук «е»)

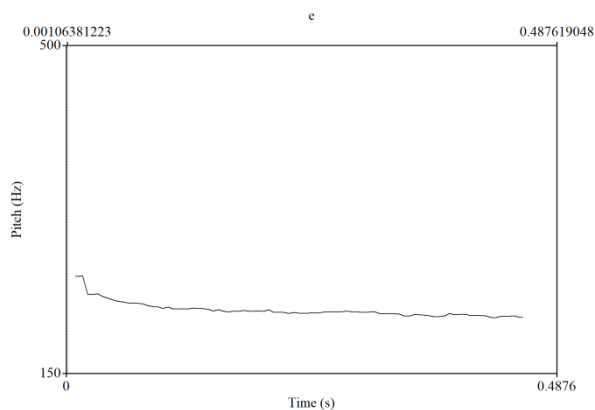


Рис. 9 Зображення висоти тону заданого сигналу (звук «е»)

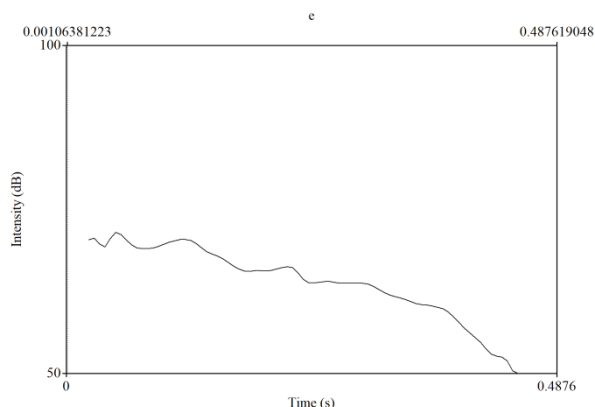


Рис. 10 Зображення інтенсивності заданого сигналу (звук «е»)

- F3_Hz (mean) = 3110.073286615095 Hz
- F4_Hz (mean) = 4256.413633226974 Hz

Як бачимо частота 1 і 3 формант є нижчою для звуку «е» ніж для звуку «а», тоді як для 2 і 4 це значення вище. Далі оцінимо висоту тону звуку.

Наведемо також мінімальну і максимальну висоту тону, а також середнє значення:

- minimum pitch: 209.40651373218557 Hz
- maximum pitch: 256.5155226612028 Hz
- mean pitch: 217.69575761434342 Hz

Для двох наведених сигналів середні значення були отримані досить близькі.

Далі розглянемо інтенсивність даного звуку.

Так як і в попередньому випадку визначимо мінімальну і максимальну інтенсивність, а також її середнє значення:

- minimum intensity: 43.85037201425328 dB
- maximum intensity: 71.53797579877471 dB
- mean-energy intensity: 66.27851868822907 dB

Як видно з отриманих результатів інтенсивність звуку «е» є дещо меншою ніж звуку «а»

Наведемо також загальний звіт щодо даного сигналу (Jitter, Shimmer і розриви голосу):

Voicing:

- Fraction of locally unvoiced frames: 5.319% (5 / 94)
- Number of voice breaks: 0
- Degree of voice breaks: 0 (0 seconds / 0 seconds)

Jitter:

- Jitter (local): 0.738%
- Jitter (local, absolute): 33.832E-6 seconds

Shimmer:

- Shimmer (local): 6.034%
- Shimmer (local, dB): 0.544 dB

Jitter і Shimmer є дещо більшими ніж в звуку «а».

Перейдемо до приголосних звуків. Слід зазначити, що вимовлялися ці звуки фактично як «ни» та «пи». Почнемо зі звуку «н» і, перш за все, наведемо його форму.

Як видно з рис.11 амплітуда значно зросла при переході на голосний «и».

Наведемо спектрограму сигналу

Очевидно, що більш затемненими є області, які також знаходяться ближче до наявного голосного звуку. Як і в попередніх випадках розглянемо форманти даного звуку.

Наведемо повну інформацію по наявних формантах:

- F1_Hz (mean) = 379.0219272651137 Hz
- F2_Hz (mean) = 2098.5695324341964 Hz
- F3_Hz (mean) = 3235.8239232376313 Hz
- F4_Hz (mean) = 4172.531650642418 Hz

Як видно з отриманих результатів частота 1 і 4 формант є нижчою для звуку «н» ніж голосних звуків, тоді як для 2 це значення вище. А от для 3 значення є більшим ніж у звуку «е», проте меншим ніж в «а».

Оцінимо висоту тону звуку.

Наведемо також мінімальну і максимальну висоту тону, а також середнє значення, так як це було зроблено вміще:

- minimum pitch: 194.62392947663398 Hz
- maximum pitch: 234.5189456896151 Hz
- mean pitch: 220.19280481973254 Hz

Очевидно, що висота тону звуку залишається на приблизно однаковому рівні. Далі розглянемо інтенсивність даного звуку.

Отримані результати:

- minimum intensity: 46.17942063923519 dB
- maximum intensity: 76.18877948259988 dB
- mean-energy intensity: 69.88919911004538 dB



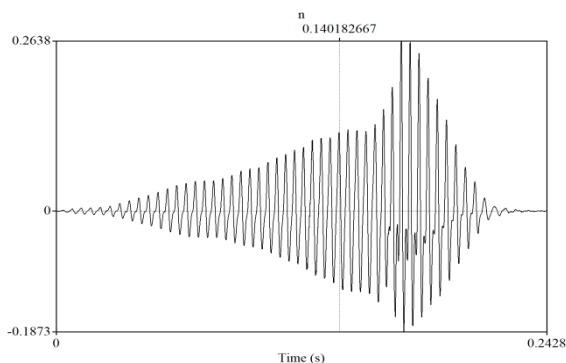


Рис. 11 Форма заданого сигналу (звук «н»)

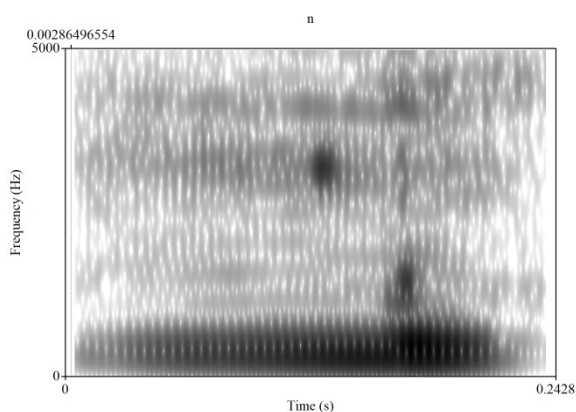


Рис. 12 Спектрограма заданого сигналу (звук «н»)

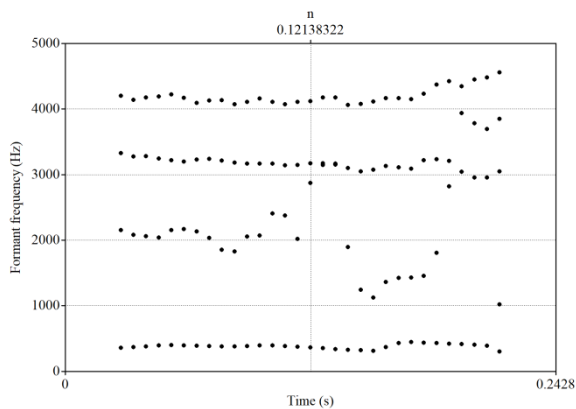


Рис. 13 Зображення формант заданого сигналу (звук «н»)

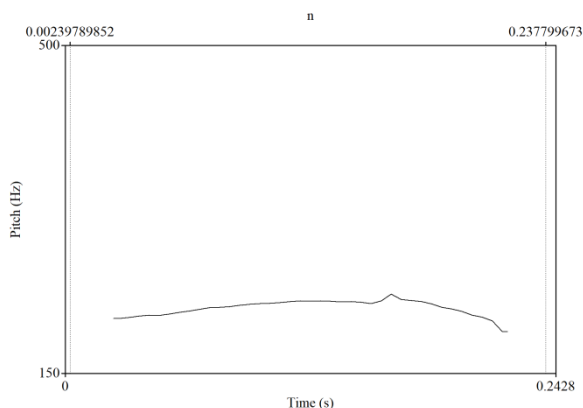


Рис. 14 Зображення висоти тону заданого сигналу (звук «н»)

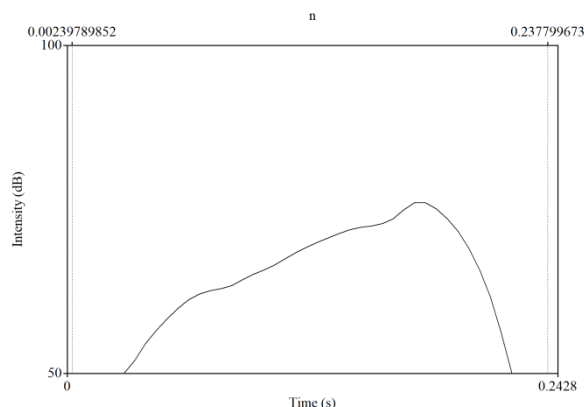


Рис. 15 Зображення інтенсивності заданого сигналу (звук «н»)

Як видно з результатів інтенсивність звуку «н» є дещо більшою, ніж звуку «е».

Загальний звіт щодо даного сигналу (Jitter, Shimmer і розриви голосу):

Voicing:

- Fraction of locally unvoiced frames: 13.333% (6 / 45)
- Number of voice breaks: 0
- Degree of voice breaks: 0 (0 seconds / 0 seconds)

Jitter:

- Jitter (local): 0.977%
- Jitter (local, absolute): 44.328E-6 seconds

Shimmer:

- Shimmer (local): 12.509%
- Shimmer (local, dB): 1.100 dB

Jitter, Shimmer та розриви голосу є дещо більшими в відсотковому відношенні ніж в попередніх випадках.

Далі розглянемо звук «п».

Як видно з рис.16 форма даного сигналу суттєво відрізняється від усіх попередніх сигналів і можна побачити звук «и» .

Як видно з рис.17 спектрограма значно відрізняється від попередніх.

Розглянемо форманти даного звука.

- F1_Hz (mean) = 402.50561853139646 Hz
- F2_Hz (mean) = 1052.8564133193981 Hz
- F3_Hz (mean) = 3031.3475089883736 Hz
- F4_Hz (mean) = 4053.7290973341655 Hz

Як видно з отриманих результатів різниця між 2 та 3 формантою становить майже 2кГц.

Оцінимо висоту тону звуку.

- minimum pitch: 207.95160056851714 Hz
- maximum pitch: 249.07005957290949 Hz



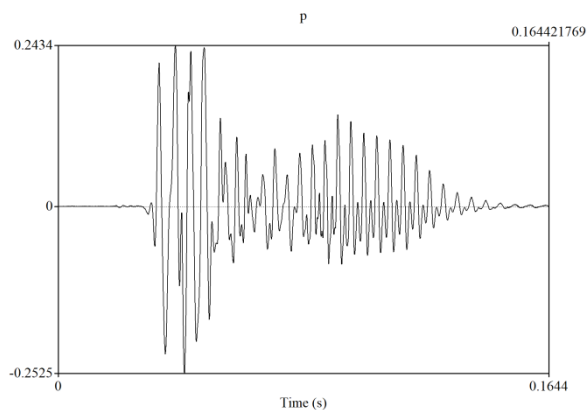


Рис. 16 Форма заданого сигналу (звук «п»)

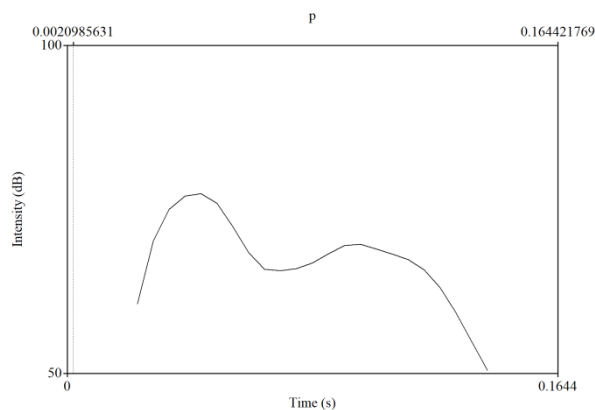


Рис. 20 Зображення інтенсивності заданого сигналу (звук «п»)

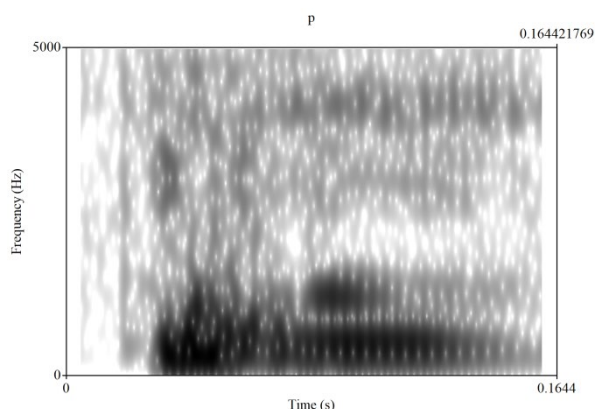


Рис. 17 Спектрограма заданого сигналу (звук «п»)

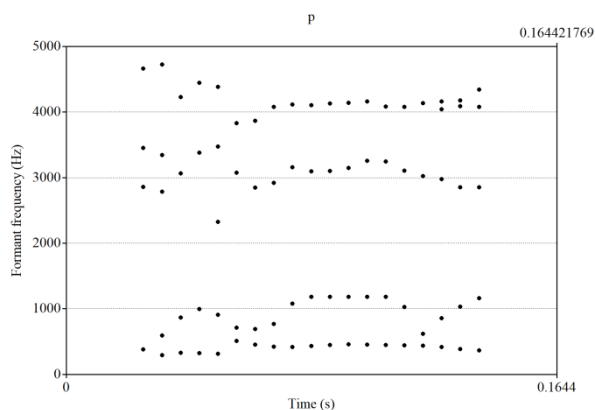


Рис. 18 Зображення формант заданого сигналу (звук «п»)

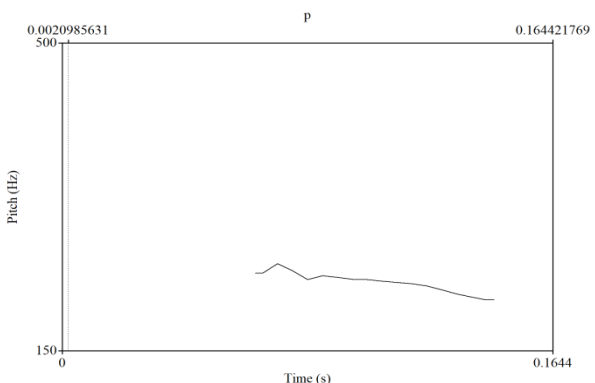


Рис. 19 Зображення висоти тону заданого сигналу (звук «п»)

- mean pitch: 228.24619872239566 Hz

Отримані результати доводять, що висота тону звуку майже не змінюється.

Розглянемо інтенсивність даного звуку.

Отримані результати:

- minimum intensity: 50.45096711599676 dB
- maximum intensity: 77.49308546521412 dB
- mean-energy intensity: 70.86791747652238 dB

Як видно з результатів середня інтенсивність звуку «п» є приблизно такою ж, як і звука «н».

Загальний звіт щодо даного сигналу (Jitter, Shimmer і розриви голосу):

Voicing:

- Fraction of locally unvoiced frames: 9.524% (2/21)
- Number of voice breaks: 0
- Degree of voice breaks: 0 (0 seconds / 0 seconds)

Jitter:

- Jitter (local): 1.787%
- Jitter (local, absolute): 78.403E-6 seconds

Shimmer:

- Shimmer (local): 15.985%
- Shimmer (local, dB): 1.659 dB

Jitter та Shimmer голосу є досить великими в порівнянні з попередніми випадками.

ВИСНОВКИ

Програмний комплекс Praat є надзвичайно корисним інструментом для аналізу і в певній мірі порівняння звуків.

Слід також зазначити, що для звуків «п» та «н» значення 1, 3 та 4 формант різняться в межах 200 Гц, але для 2 форманти різниця більш суттєва (трохи більше 1 кГц). А от для звуків «а» та «е» значення 1 форманти є більшим, аніж для розглянутих приголосних звуків.



Середня інтенсивність виявилась найменшою для звуку «е», та найбільшою для звуку «а». Найбільш високою інтенсивністю характеризуються голосні звуки, менш інтенсивними є приголосні [10]. Отримані результати ж вказують на наявність звуку «и» при вимові приголосних, тому їх середня інтенсивність знаходиться приблизно на одному рівні.

Також зазначимо, що Jitter і Shimmer голосу в відсотковому відношенні було більшим, для приголосних звуків.

Як видно з отриманих результатів, при аналізі звуків мови Praat дозволяє автоматизувати оцінювання головних параметрів, що визначають звуки мови — це формантний склад звуків, інформація про який потрібна для спеціалістів із побудови систем автоматичного синтезу та розпізнавання мови.

Praat також дозволяє оцінювати низку інших параметрів, інформація про які є корисною для медичних спеціалістів (зокрема фоніатрів) — це такі параметри як зміна інтенсивності у часі, виявлення переривань голосу, параметри Shimmer та Jitter. Перевагами програми є її доступність, легкість у використанні, завдяки простому інтерфейсу і можливість скористатися довідкою, а також можливість створювати і експортувати високоякісні рисунки, що значно процес обробки отриманих даних.

Підсумовуючи наведені вище міркування, можна зробити висновок, що використання програмного комплексу Praat для аналізу мови шляхом порівняння обраних звуків є доцільним.

Надійшла до редакції 02 квітня 2021 р.

ПЕРЕЛІК ПОСИЛАНЬ

- [1] AN URSS. Instytut movoznavstva, z. r. (1970—1980). Slovyk ukrayins'koyi movy v 11 tomakh [Dictionary of the Ukrainian language in 11 volumes]. Kyiv: Naukova dumka.
- [2] Cambridge University Press, 2. *Cambridge Dictionary*. [Online]. Available: <https://dictionary.cambridge.org/ru/%D1%81%D0%BB%D0%BE%D0%B2%D0%B0%D1%80%D1%8C/%D0%B0%D0%BD%D0%B3%D0%BB%D0%B8%D0%B9%D1%81%D0%BA%D0%B8%D0%B9/speech>
- [3] Shydlovs'ka, T. A. DEYAKI VIDOMOSTI PRO HOLOS. [SOME INFORMATION ABOUT VOICE] Otrymano z DU «Instytut otolarynholohiyi im. prof. O.S. Kolomyichenka NAMN Ukrainy», Laboratoriya profesynykh porushen' holosu ta slukhu: 2011. URL: http://www.audiovoice.kiev.ua/interest_01.shtml
- [4] Paul Boersma and David Weenink, P. S. (б.д.). *Praat: doing phonetics by computer*. URL: <https://www.fon.hum.uva.nl/praat/>
- [5] University Purdue, P. R., *Fourier Analysis and the Speech Spectrogram*. URL: https://www.projectrhea.org/rhea/index.php/Speech_Spectrogram
- [6] Pascal van Lieshout, P. (2003). *PRAAT Short Tutorial*.
- [7] Karaman S. O., Karaman O. V., Pliushch M. Ya.. *Suchasna ukrayins'ka literaturna mova. Navchal'nyy posibnyk [Modern Ukrainian literary language. Tutorial]*. Litera LTD, 2011
- [8] *Muzychnyy ton. Huchnist' i vysota tonu [Musical tone. Volume and pitch]*. URL: <https://sites.google.com/site/kolivannihvilijk/home/muzicnij-ton-gucnist-i-visota-tonu>
- [9] Mireia Farrús, J. H. Jitter and shimmer measurements for speaker recognition. *INTERSPEECH, 8th Annual Conference of the International Speech Communication Association*, pp. 778-781. Antwerp, Belgium, 2007.
- [10] Yu. L. Trofimov. *Inzhenerna psikhoholohiya: Pidruchnyk [Engineering Psychology: Textbook]*. Kyiv, Lybid, 2002



Analysis of Speech Sounds In the Praat Program

A. V. Pareniuk, ORCID [0000-0002-8882-5976](https://orcid.org/0000-0002-8882-5976)

Department of Acoustics and Acoustoelectronics ames.kpi.ua

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", ROR [00syn5v21](https://ror.org/00syn5v21)
Kyiv, Ukraine

Abstract—The article identifies the possibility of using the Praat program for comprehensive and complete analysis of speech sounds. It also indicates the range of functions of the software package and the perspective of using the program for comparing sounds.

Praat is a program for speech analysis and synthesis written by Paul Boersma and David Veninko at the Department of Phonetics, University of Amsterdam.

Praat is a very flexible tool for speech analysis. It offers a wide range of standard and non-standard procedures. The functions of the program that were used in the work: spectrograms, pitch analysis, formant analysis, intensity analysis, jitter, shimmer, voice breaks. There are at least three ways to get sound in the program: sound recording, read sound from the disc, create sound from formulas. In this article, the first method was used and the vowel sounds "a" and "e", as well as consonants "n" and "p" were considered.

With the help of automated algorithms, a number of important data were quickly obtained, such as the average value of the formant on the selected interval, the maximum, minimum and average value of pitch, maximum, minimum and average intensity, the overall signal report (jitter, shimmer, voice breaks). Also, this program can create high quality images that can be exported and inserted into scientific articles or personal research. Due to this fact, the images of not only spectrograms and waveforms, but also the images of formants and intensity was presented in the work.

Based on the obtained results, the selected sounds were compared. As a result, a conclusion was made about the feasibility and effectiveness of using the software package Praat in the analysis of speech sounds.

It was also noted the ease of use of this software package, as well as the importance of using the program in research in various fields of human activity.

Keywords — *spectrogram; formant; pitch of the sound; intensity; Jitter; Shimmer; voice breaks; sound; Praat software package; frequency.*