

ІНВАРІАНТНІ АКУСТИЧНІ ХАРАКТЕРИСТИКИ УКРАЇНСЬКИХ ГОЛОСНИХ ЗВУКІВ

На основі інваріантно-акустичного аналізу мовлення отримані постійні спектральні характеристики українських голосних для різних способів вимови: звичайного мовлення, шепоту та змінного тону. Показано, що акустичними інваріантами звуків мовлення є відношення між частотами головних формант, але не їх абсолютні значення. Продemonстровано, що акустичні інваріанти українського звука [i] та англійського [I] однакові.

Ключові слова: акустична фонетика, розпізнавання мовлення, звуки мовлення, акустичні параметри, акустичні інваріанти.

Метою роботи є визначення акустичних інваріантів для голосних звуків української мови. **Актуальність** теми зумовлена, з одного боку, недостатнім використанням акустичного підходу до аналізу мовлення та відсутністю чіткого розмежування інваріантних та варіантних параметрів людського мовлення, а з іншого – потребою створення розпізнавачів мовлення нового покоління, що базуються на використанні незалежних від мовця характеристик звуків мовлення, а також необхідністю укладання атласу звуків української мови. **Предметом** дослідження є акустичні характеристики українських голосних звуків, **об'єктом** – звуки мовлення.

Як відомо, мова є комунікативною основою людської діяльності: адже для обміну інформацією ми використовуємо мову – здебільшого таку, що має звукову основу. А звуки мовлення – ті першотвірні цеглинки, з яких складається мова, – досліджує фонетика, спираючись при цьому на фізичну інтерпретацію відповідних акустичних явищ і процесів. Тому фонетику в певному сенсі можна назвати базовим розділом мовознавства. Крім того, фонетика є особливою галуззю науки: вона лежить на стику фізики та лінгвістики. З одного боку, за міжнародною класифікацією PACS (Physics та Astronomy Classification Scheme), ця наука є підрозділом акустики (її номер 43.70 Hs), а отже, фізики. І користується фонетика фізичними методами дослідження. На цьому наголошував, зокрема, і М.С.Трубецької: “Наука про звуки мовного акту, яка має справу з конкретними фізичними явищами, повинна користуватися природничими методами...” ([1]). З іншого боку, у фонетики суто мовознавчий об'єкт вивчення: це звуки мовлення. А інформація про звуки мовлення, одержана в результаті цих досліджень, знаходить своє застосування не лише в мовознавстві – а це вироблення правил написання запозичених слів ([2]), вивчення іноземних мов ([3]) та створення транслітераційних систем ([4]) для міжнародних і

національних стандартів, – а й у різноманітних галузях науки й техніки, зокрема в моделюванні та розпізнаванні мовлення ([5-14]), включаючи таємні комп'ютерні технології.

Створення акустичної бази даних для комп'ютерного моделювання мовлення (генерації звукових сигналів) та його розпізнавання (ідентифікації звукових сигналів) вимагає докладної інформації про фізичну суть звуків мовлення та їх нелінійну взаємодію. Ця інформація використовується на стадії попереднього аналізу даних. Глибоке розуміння процесів, які зумовлюють формування фонем, відкриває можливість створення пристрою, здатного розпізнати будь-який потік мовлення незалежно від віку, статі та емоційного стану мовця – такого собі “комп'ютерного секретаря”. З іншого боку, належне розуміння впливу психологічних чинників на звукові параметри дозволило б створити надійні детектори брехні, визначники захворювань і звукові замки з вибірковою реакцією на певний емоційний стан. Цілком імовірно, колись ми станемо спроможними точно змодельовати вимову будь-якої особи та відтворити голоси відомих співаків минулого. До того ж цікавою є можливість доповнення комп'ютерних словників звуковим супроводом. Усі ці завдання тією чи іншою мірою торкаються багатьох науково-технічних проблем – зокрема, створення штучного інтелекту.

На сьогодні існує багато систем автоматичного розпізнавання мовлення, які можуть задовольнити більш-менш вишукані вимоги. Втім, сучасні розпізнавачі, як зізнаються відомі спеціалісти Вейн Лі [13] та Тарас Вінцюк [14], поки що не можуть повністю розв'язати проблеми шумів, навчання, статі, емоційного стану тощо. Зокрема, відповідні комп'ютерні програми *Dragon Dictate* та *Kurzweil* дуже чутливі до цих чинників, вимагають тривалого навчання, роблять під час розпізнавання багато помилок та ще й потребують значного обсягу комп'ютерної пам'яті. Отже, з одного боку, штучний розум досяг такого високого рівня, що “Deer Blue” виграє в чемпіона світу з шахів – а тут комп'ютери досі не можуть упоратися з “проблемами диктора” (розуміння іншого мовця), завад (виділення корисного сигналу з шумів) чи емоційного стану. Будь-яка дворняжка легко розпізнає свою кличку – навіть якщо чує цю людину вперше – а ще чудово розрізняє голоси і навіть інтонацію. І все це на гамірній вулиці! Слід визнати, що Жучка поки що може дати велику фору відповідним “розумним” пристроям, на виготовлення яких кидаються величезні кошти... А справа в тому, що в таких пристроях закладається надмірна

кількість параметрів розпізнавання (як акустичних, так і артикуляційних), унаслідок чого інваріантні та суб'єктивні характеристики звуків мовлення виявляються переплутаними. Як бачимо, означена проблема все ще вимагає подальшого дослідження. Тому такою важливою видається узгоджена фізична теорія утворення звуків мовлення та одержання надійних інваріантних характеристик звукових спектрів.

Вагомим чинником, який зумовлює подальший поступ автоматичного розпізнавання мовлення, є сьогодиншній стан фонетичної науки. Дотеперішні дослідження людського мовлення [5-22] базувалися переважно на артикуляції. Порівняно з акустичними властивостями звуків мовлення, їх артикуляція легше піддається дослідженню та класифікації і не так істотно залежить від програмного забезпечення та приладдя – тому й вивчена набагато глибше та детальніше. Про це свідчить і монографія відомого американського фонетиста К.Стівенса “Акустична фонетика” [5], де акустичні характеристики звуків мовлення виводяться саме з артикуляційних моделей. Як більш розроблений, артикуляційний підхід широко застосовується і в сучасних системах розпізнавання мовлення, що переважно базуються на так званій генеративній моделі [14], де розпізнаваний сигнал порівнюється з опорними сигналами, які генеруються комп'ютером на основі даних про артикуляцію. То й не дивно, що такий підхід не дає змоги відділити незалежні від мовця характеристики звука від залежних, які відображають індивідуальність людини. До того ж, як зазначено в [21], дотеперішні моделі мовлення не можуть достеменно відтворити всі особливості його механізму.

У природній же мовленнєвій комунікації вирішальне значення має не артикуляційний, а якраз акустичний механізм (ми тут не будемо торкатися психологічних аспектів сприйняття мовлення). Саме за допомогою акустичного аналізу живі істоти інтерпретують звукову хвилю, яка змушує коливатися їхні барабанні перетинки. Нам зовсім не обов'язково знати всі деталі артикуляційного творення звуків, щоб зрозуміти співрозмовника. Якби для розуміння мовця нам доводилося артикулювати почуті звуки мовлення, то ми просто не змогли б сприймати одночасне мовлення кількох осіб. Діти вчаться говорити на основі почутих звуків, а знання артикуляції відіграє лише допоміжну роль. Було доведено теоретично та підтверджено експериментально, що мовці прагнуть відтворити відстані між формантами відповідно до акустичної цілі (див. [11]). Саме тому глухі від народження люди є німи: їм бракує звукової інформації для проведення акустичного аналізу мовлення в їхньому мозку. А навчити глуху люди-

ну говорити бездоганно – використовуючи артикуляційний підхід – є завданням неможливим. Саме завдяки акустичному аналізу вчаться говорити й папуги – адже їхній артикуляторний апарат зовсім не такий, як у *Homo sapiens*.

Кожен звук мовлення характеризується головними (постійними, перманентними) акустичними параметрами, які притаманні даному звуку незалежно від мовця, та побічними (змінними) сумірниками, які характеризують індивідуальність мовця чи даної вимови (це відображається в особливостях швидкості мовлення, тональності, тембру, звукової взаємодії і т. ін.). Перший клас параметрів описується відношенням характерних частот – у цьому неважко переконатися, якщо згадати, що відтворення записаного мовлення на іншій швидкості не змінює якості звуків мовлення (наприклад, [а] не перетвориться при цьому на [о]). Другий клас параметрів визначається як абсолютними величинами – такими, як частота основного тону (ЧОТ) та резонансні частоти звукового тракту мовця, – так і відносними – як, наприклад, відносна інтенсивність обертонів чи шумів. Отже, й форманти звуків мовлення розпадаються на два класи: головні (постійні, перманентні) та побічні (змінні). Головні форманти характеризуються інваріантним відношенням їх частот, незмінним для даного звука мовлення. І завданням інваріантно-акустичного аналізу звуків мовлення є визначення цих величин.

Як же, проводячи акустичні експерименти, відрізнити найнижчу головну форманту від нижчих гармонік основного тону, зумовлених голосовою складовою? Є два шляхи “виключення” ефектів голосу: це шепіт і спів. У шепоті голосова складова взагалі відсутня, а при співі (або інтонуванні мовлення) ЧОТ – а з нею і частоти її гармонік – змінюються. Незмінними залишаються тільки формантові частоти (ФЧ), які, як відомо, зумовлені резонансом.

Отже, інваріантно-акустичний аналіз звуків мовлення – це якісно новий метод акустичної фонетики, який дозволяє виявити і прокласифікувати важливі акустичні параметри. Цей підхід є важливим чинником ефективного розвитку сучасної експериментальної фонетики та автоматичного розпізнавання мовлення.

У цій розвідці ми дослідимо акустичні характеристики мовленнєвих хвилеформ, які відіграють вирішальну роль у сприйнятті та розпізнаванні мовлення. Як уже було зазначено, кожен звук людського мовлення містить акустичну інформацію як про сам звук, так і про мовця. Для того, щоб відділити одні характеристики від інших, тут застосована методика інваріантно-акустичного аналізу мовлення. Ми з'ясуємо також фізичну картину утворення звуків мовлення та проаналізу-

емо спектральні характеристики українських голосних для різних способів вимови: звичайного мовлення, шепоту та змінного тону (співу). Буде показано, що найнижчі фонемні частоти, зумовлені коливаннями голосових зв'язок і резонансом Гельмгольца, не можуть бути пов'язані з головними формантами – за винятком деяких випадків мовлення високим тоном, коли основна частота дорівнює частоті першої форманти: $F_0 = F1$. Якщо ЧОТ настільки висока, що в резонаторі вміщується дві або більше хвиль кратної частоти (обертонів), виникають додаткові високочастотні резонанси – а з ними додаткові форманти на частотах, пропорційних частотам головних формант. Найважливіший висновок полягає в тому, що єдиним фонемним інваріантом є відношення постійних ФЧ. Цей аналіз доповнений комп'ютерним синтезом деяких звуків. Показано, що акустичні інваріанти українського звука [і] та англійського [I] однакові. Якщо звук “змішаний” з іншим, у ньому виникають додаткові максимуми на частотах, відношення яких відповідає домішаному звуку. Отримані результати можуть бути корисними для спеціалістів у галузі експериментальної фонетики та моделювання мовлення.

Нагадаємо, що в мовленнєвому тракті (МТ) людини можливі два типи резонансу. По-перше, це низькочастотний резонанс Гельмгольца (РГ), який виникає в об'ємі зі звуженням з одного боку (відкритою трубою меншого діаметра) – тобто відповідний резонатор має пляшкоподібну конфігурацію. По-друге, це трубний резонанс (ТР), або резонанс у трубі, де труба може бути закритою з обох кінців чи з одного (див. [5]). Для конфігурації, що складається з великого об'єму (який зумовлює акустичну податливість) та звуження (акустична маса), де розміри об'єму та звуження малі порівняно з довжиною хвилі $\lambda_H = c/f_H$ (c – швидкість звукової хвилі), власна частота Гельмгольца становить

$$f_H = [c/(2\pi)] [A/(Vl)]^{1/2}, \quad (1)$$

де A та l – площа поперечного перерізу та довжина вузької труби відповідно, а V – об'єм широкої труби (подробити див. у [5, с. 142]). Цей резонанс може існувати при достатньо низьких частотах (приблизно до 500 Hz), де вищі гармоніки не задіяні.

Власні (вільні) частоти напівзакритої труби довжиною l зі сталим поперечним перерізом зумовлюються чвертьхвильовим резонансом і становлять

$$f_n = [(2n - 1)/4](c/l), \quad (2)$$

де $n = 1, 2, 3, \dots$ – номер власної частоти (подробити див. у [5, с. 138-139]).

Якщо труба замкнута чи відкрита з обох кінців, виникає півхвильовий резонанс, і

$$f_n = [(n - 1)/2](c/l). \quad (3)$$

Якщо ЧОТ перевищує частоту РГ (або навіть якщо трохи менша за неї), то мають місце тільки (2) та (3), при цьому параметр n пробігає вищі значення.

Слід розрізняти власні частоти голосового тракту, які мають фіксовані значення для даної його конфігурації у даної людини, та ФЧ, які визначають даний звук мовлення (ця відмінність не відзначена в [5]). Перші величини, безумовно, залежать від особистості мовця. Якщо частота якої-небудь парціальної гармоніки у спектрі джерела (включаючи ЧОТ) наближається до вільної частоти МТ, її амплітуда зростає. Це спричиняє утворення бажаного (чи випадкового) звука мовлення. Якби нам вдалося визначити спосіб, у який максимумами кожного звука мовлення пов'язані з незалежними від мовця чинниками, це допомогло б істотно наблизитися до успішного розв'язання проблеми розпізнавання мовлення.

При цьому слід чітко уявляти, що таке форманта – і дати таке її означення, що точно та адекватно відображає суть поняття. Ось посібник [22, с. 42] означає форманти як “зони посилення частот”. Таке формулювання не має фізичного змісту. По-перше, не треба плутати форманту із зоною чи частотою – це різні поняття. По-друге, посилюється те, що має силу: вітер, обійми тощо – а частоти можуть зростати. Насправді ж у резонансній області відбувається зростання амплітуд звукових коливань – а ніяк не частот. До того ж, як ми побачимо нижче, деякі гармоніки, що визначають даний звук мовлення, не обов'язково інтенсивніші за інші (це трапляється за РГ, а також коли один звук “змішаний” з іншим, або якщо звук перебуває під впливом сусіднього). Тому будемо розглядати форманту як моду звукової хвилі (обертон), що потрапляє в область резонансу і дає внесок в утворення даного звука мовлення ([23, с. 21; 24]).

А тепер проаналізуємо особливості РГ у звуковій хвилі, щоб з'ясувати, якою мірою він може стосуватися перманентних характеристик звуків мовлення.

1. Коливання тиску повітря, що надходить у людське вухо чи мембрану мікрофона, описуються добре відомим виразом

$$p_r(f) = [i\bar{p}/(2r)] U_r(f) \exp(-2\pi i f r / c), \quad (4)$$

де i – уявна одиниця, f – частота, ρ – густина повітря, c – швидкість звуку в повітрі, r – відстань від рота, p_r – звуковий тиск на відстані r , U_r – об'ємна швидкість на відстані r (див., напр., [5, с. 127-128]). Із (4) видно, що низькочастотне збудження спричиняє меншу амплітуду тиску, ніж високочастотне, – тобто моди з вищими частотами є виразнішими. Це вказує на те, що найважливіші акустичні характеристики звуків мовлення містяться, найімовірніше, в діапазоні середніх і високих частот.

2. Унаслідок великої дисипації енергії в низькочастотній області, резонанс Гельмгольца у головному тракті зазвичай характеризується значними ширинами та не дуже великими амплітудами. Ці втрати енергії зумовлені імпедансом (повним опором) стінок МТ ([5, с.157,193; 18; 19]), в'язкісними та тепловими втратами ([15; 6], див. також [5, с. 160-161]), розкриттям глотки ([5, с. 165-166]) чи звуженням мовленнєвої порожнини ([5, с.534]), а також нелінійним акустичним опором (5, с. 163-164]). Відтак зменшується можливість виявлення відповідної амплітуди – а з нею й роль такої моди в утворенні звука. Для більшості конфігурацій МТ у мовленні шепотом акустичні втрати внаслідок широкого глоткового розкриття такі великі, що відбувається критичне загасання РГ, і відповідний спектральний пік повністю згладжується (див. [5, с. 165, 171] та [16; 17]). Проте звуки [i], [и], [y], [л], [м], [н], [с] та інші, в яких нижні форманти, спричинені РГ, не зникають зовсім: їх можна почути й ідентифікувати навіть у шепоті.

3. Поверхня язика, щік і гортані не є твердою, а отже, виникає реактивний опір маси стінок, який зумовлює істотну зміну найнижчої власної частоти МТ. У працях [18; 19] було знайдено, що ця поправка становить близько 180 Hz для дорослого чоловіка та 190 Hz для дорослої жінки. Через це низькі частоти найменше контролюються артикуляційними рухами. У праці [5, с. 160] відзначається, що частота 250 Hz лише наполовину чутлива до зміни конфігурації МТ. Це знову вказує на те, що частоти Гельмгольца навряд чи відіграють важливу роль у формуванні суттєвих особливостей звуків мовлення.

4. Експерименти зі сприйняття голосних [8; 9] показали, що сукупність найнижчих гармонік інтерпретується як єдиний максимум, якщо вони розташовані достатньо близько до ЧОТ – приблизно не далі ніж 300 Hz у діапазоні нижче 500 Hz.

Втім, людське вухо може відчувати значно меншу різницю частот. Наприклад, музиканти та співаки спроможні розрізняти інтервали в чверть тону, що відповідає кільком герцам (чи навіть менше) у діапазоні частот мовлення. Отже, наш мозок, аналізуючи звукову хвилю, просто не бере до уваги деякі найнижчі частоти. ЧОТ ігнорувати не можна, тому що висота тону важлива в інтонуванні мовлення та співі. Тому виходить, що саме частоти Гельмгольца сприймаються нашим мозком як найменш важливі для розрізнення звуків.

5. Згадаймо, що ЧОТ, яка визначає висоту тону, найнижча в звуковій хвилі. Нижче неї не може бути ніяких максимумів. Утім, ця частота легко досягає значення 500 Hz у звичайному мовленні та 1000 Hz і вище під час співу – тоді як РГ зумовлює частоти не вище приблизно 500 Hz. Але під час мовлення та співу високим тоном звуку [i],

[и], [y], [л], [м], [н], [с], найнижчі форманти яких, як зазначено в [22, с.57], мали би міститися в діапазоні 200-400 Hz, не вироджуються. Це говорить про те, що низькі частоти не дуже істотні для утворення звуків мовлення.

Підсумовуючи сказане, слід зазначити, що ефекти Гельмгольца невиразні або цілком відсутні у багатьох способах мовлення і що відповідний резонанс не може відігравати помітної ролі в ідентифікації звуків мовлення. Таким чином, РГ – за умови достатньо низької ЧОТ – може давати внесок лише у побічні характеристики, що відображають індивідуальність мовця (такі, як тембр голосу) та ефекти, пов'язані зі взаємодією звуків.

Щоб відділити інваріантні акустичні характеристики від варіативних, згадаймо той факт, що відтворення магнітофонного запису на підвищеній швидкості ("голос Буратіно") не змінює один звук мовлення на інший (тобто, наприклад, [а] залишається звуком [а], не перетворюючись на [о] чи на [і]) Таке відтворення еквівалентне збудженню вищих гармонік основної хвилі, яка потрапляє в ТР (нагадаємо, що РГ для обертонів неможливий). Іншими словами, якщо кожен ФЧ звука мовлення пропорційно змінити, то якісні ознаки цього звука зберігаються. Отже, інваріантом кожного звука мовлення може бути лише відносна величина, зумовлена ТР – а саме відношення між частотами його формант. Це логічно пояснює результати [12], де спостерігалася кореляція між F_0 і формантовими частотами голосних звуків, і [25], де мікродинамічна поведінка французьких, англійських і чеських голосних виявилася незалежною від ЧОТ. Звертали увагу на мінливість абсолютних значень ФЧ і В. Зу та Р. Шварц ([13, с. 165]). У термінах відношень ФЧ ми можемо пояснити й той факт, що різні люди, маючи різні довжини голосового шляху та відповідно різні ФЧ, здатні вимовляти ті ж самі звуки.

Тепер зрозуміло, чому деякі музичні інтервали приємні для слуху (це питання поставив Р.Фейнман у [26, с. 440]): просто такі відношення частот природно відображають сутність звуків людського мовлення.

Щоб дістати відношення, потрібно два параметри. Тому кожен звук мовлення (крім носових) повинен мати дві незалежні головні ФЧ (які частіше відповідають резонансам у передній і задній порожнинах), а носові мають три таких частоти (додаткова ФЧ зумовлена носовою порожниною). Це твердження узгоджується з експериментальними даними [10], які свідчать, що для задовільного сприйняття неносових голосних потрібно лише дві форманти.

Експериментальна частина цієї роботи полягає не лише в тому, щоб зібрати певний масив акустичних даних, але й перевірити наші теоре-

тичні припущення про те, що інваріантні параметри звуків мовлення є відносними величинами та що вони не можуть зумовлюватися РГ. Щоб провести інваріантно-акустичний аналіз мовлення, ми будемо шукати такі постійні форманти за змінних умов: це різні мовці, різні способи вимови та застосування альтернативних процедур хвильового аналізу. Також обчислимо відповідні відношення.

У наших експериментах було задіяно 4 українських мовці: дві жінки (19 та 26 років) і два чоловіки (по 34 роки). Вони вимовляли (повторюючи тричі) окремі звуки нормальним тоном, пошепки та змінним тоном (угору та вниз), а також слова та комбінації слів: *хата, ха-ха, гадати, ходити, сидіти, лежати, я піду гуляти, бики, боки, якби, вийти, увійти, за Сибіром сонце сходить, зілля, шиття, жіночі хитрощі, докір, здійсмати, клепки, легко, хтось, не треба, це буде, як з кроликами* – у нормальному тоні та пошепки. Така процедура дає змогу відібрати характеристики, незалежні від способу вимови. Ми не ставили за мету зібрати великий масив акустичних даних від великої кількості мовців. Ключова ідея полягає у використанні різних варіантів вимови: 1) нормального мовлення, 2) шепоту, 3) зростального тону, 4) спадного тону. Якщо якийсь акустичний параметр присутній у всіх варіантах вимови, його можна розглядати як можливий інваріант звука мовлення. У цьому сенсі акустичний матеріал,

одержаний від одного мовця, є значно більш репрезентативним, ніж комбінація “нормальна вимова чотирьох мовців”.

Звукові сигнали були записані в ізольованій кімнаті за допомогою однонапрямого динамічного мікрофона Tech TDM-204 на відстані 5-10 см від губ суб'єкта мовлення, використовуючи комп'ютерну програму Sound Forge 4.0. Акустичний аналіз проводився за допомогою програм WaveLab 2.1 і CoolEdit 95. ЧОТ та ФЧ вимірювалися за допомогою: 1) осцилограми, 2) спектрограми тривимірного частотного аналізу програми WaveLab 2.1, 3) двовимірної спектрограми програми CoolEdit 95. Потім отримані результати порівнювалися та усереднювалися. Статистичні похибки зумовлюються людським фактором, точністю хвильового аналізу програм, кроком точності спектрограми та розпливанням спектрів.

У цій статті ми зосередимося на українських голосних [a], [o], [y], [и], [і], [е].

У табл. 1 наведені інваріантні частотні характеристики українських голосних: середні значення першої головної ФЧ (*Fp1*) та другої головної ФЧ (*Fp2*) разом із відповідними резонансними областями, а також відношення між цими частотами. Мовець 1 жіночої статі, вік 26 років, зріст 165 см; мовець 2 жіночої статі, вік 19 років, зріст 168 см; мовець 3 чоловічої статі, вік 34 роки, зріст 176 см; мовець 4 чоловічої статі, вік 34 роки, зріст 175 см.

Таблиця 1.

Інваріантні акустичні характеристики українських голосних

Звук	Мовець	Середнє значення <i>Fp1</i> та область резонансу, Hz	Середнє значення <i>Fp2</i> та область резонансу, Hz	Відношення <i>Fp2/Fp1</i>
[a]	1	830 (760–900)	1100 (1050–1150)	4/3 (1.33 ± 0.04)
	2	830 (800–850)	1100 (1000–1200)	4/3 (1.33 ± 0.04)
	3	820 (750–880)	1100 (1000–1200)	4/3 (1.34 ± 0.04)
	4	820 (750–900)	1100 (1000–1200)	4/3 (1.33 ± 0.04)
[o]	1	530 (480–580)	790 (710–880)	3/2 (1.49 ± 0.15)
	2	550 (510–700)	820 (720–1000)	3/2 (1.49 ± 0.15)
	3	530 (470–680)	790 (700–880)	3/2 (1.49 ± 0.15)
	4	530 (470–700)	790 (700–1100)	3/2 (1.49 ± 0.15)
[y]	1	3600 (3400–4000)	5800 (5400–6100)	5/3 (1.61 ± 0.08)
	2	4100 (3800–4300)	7000 (6400–7600)	5/3 (1.71 ± 0.06)
	3	3700 (3400–4100)	6100 (5500–6500)	5/3 (1.65 ± 0.09)
	4	3800 (3600–4000)	6400 (6200–6600)	5/3 (1.68 ± 0.04)
[и]	1	1800 (1300–2000)	2100 (1800–2400)	6/5 (1.17 ± 0.07)
	2	1800 (1600–2100)	2200 (1900–2600)	6/5 (1.22 ± 0.06)
	3	1800 (1400–2000)	2200 (1800–2700)	6/5 (1.22 ± 0.06)
	4	1800 (1500–2100)	2200 (1900–2500)	6/5 (1.22 ± 0.06)
[і]	1	2300 (2000–3000)	2900 (2500–3800)	5/4 (1.26 ± 0.08)
	2	2400 (2000–3500)	3000 (2500–4400)	5/4 (1.25 ± 0.07)
	3	2500 (2100–3400)	3100 (2700–4200)	5/4 (1.24 ± 0.08)
	4	2400 (2100–3400)	3000 (2500–4200)	5/4 (1.25 ± 0.07)
[е]	1	700 (610–780)	2100 (1600–2400)	3 (3.00 ± 0.30)
	2	700 (580–950)	2100 (1800–2800)	3 (3.00 ± 0.30)
	3	700 (550–930)	2100 (1700–2500)	3 (3.00 ± 0.30)
	4	700 (520–920)	2100 (1600–2600)	3 (3.00 ± 0.30)

Кількість гармонік найменша в [i] та найбільша в [a] та [o]. Найвиразніший резонанс (контраст амплітуд) проявлявся в [a] та [o], тоді як звук [и] виявився “найневизначнішим”.

За нормального мовлення у спектрі звука [и] та меншою мірою [i] спостерігалися значні амплітуди в діапазоні 200-500 Hz, зумовлені РГ перших

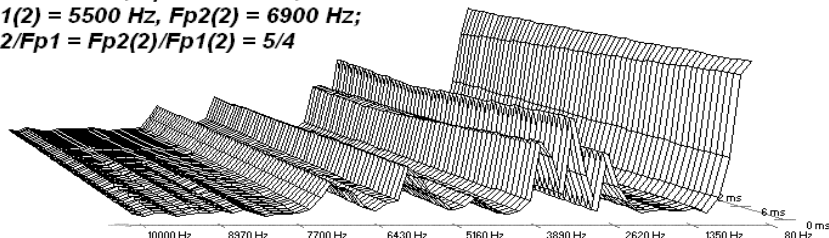
гармонік (див. мал. 1-3). Але ці максимуми зникають за підвищення ЧОТ, коли частота першого обертона достатньою мірою перевищує частоту Гельмгольца: $f_1 = 2F_0 > f_H$, і стають у край невидимими в шепоті. Це підтверджує наш висновок про те, що РГ не може давати перманентні, інваріантні характеристики звуків мовлення.

[i] вголос (ч):

$Fp1 = 2400 \text{ Hz}$, $Fp2 = 3000 \text{ Hz}$;

$Fp1(2) = 5500 \text{ Hz}$, $Fp2(2) = 6900 \text{ Hz}$;

$Fp2/Fp1 = Fp2(2)/Fp1(2) = 5/4$

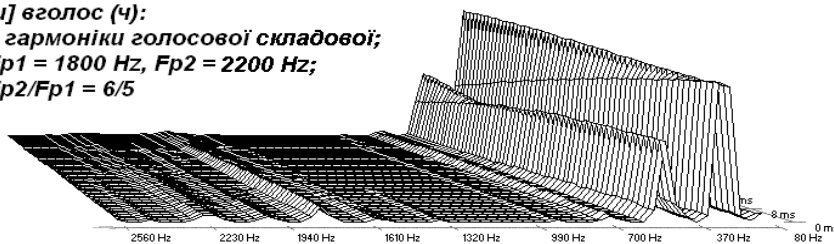


[и] вголос (ч):

2 гармоніки голосової складової;

$Fp1 = 1800 \text{ Hz}$, $Fp2 = 2200 \text{ Hz}$;

$Fp2/Fp1 = 6/5$

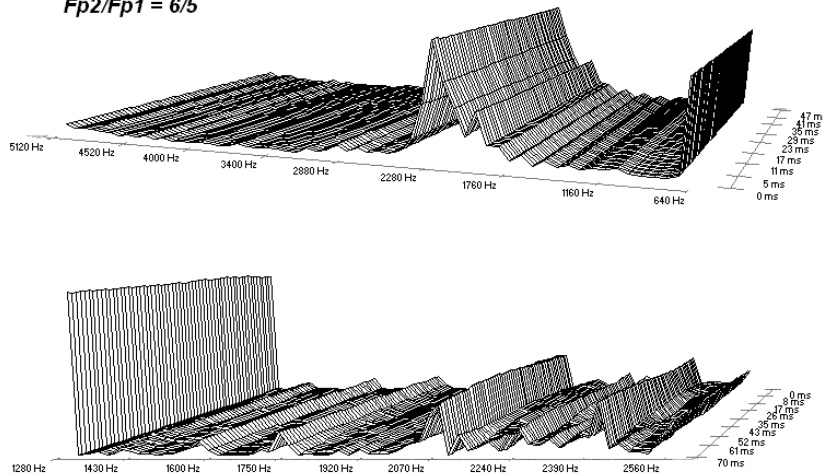


Мал. 1

Звук [и] у звичайному тоні (ч):

$Fp1 = 2000 \text{ Hz}$, $Fp2 = 2400 \text{ Hz}$,

$Fp2/Fp1 = 6/5$

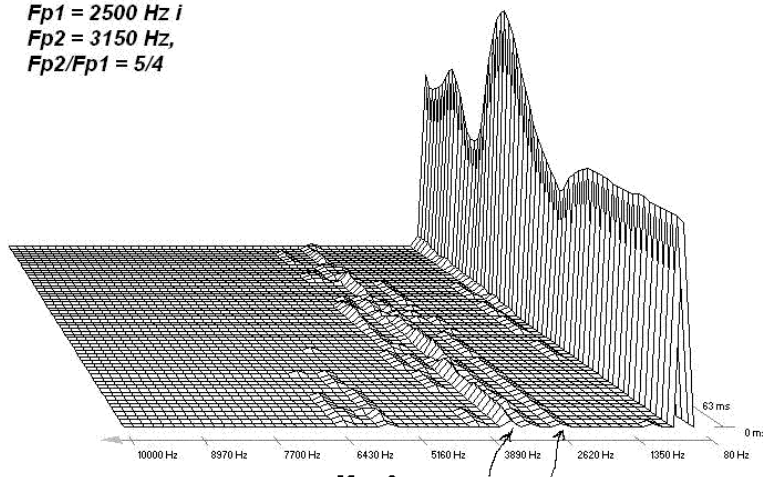


Мал. 2

Для порівняння ми дослідили форманти, зумовлені ТР у різних акустичних реалізаціях англійської фонем /i/ ([i:] та [I]), використовуючи англійський навчальний аудіокурс “Kernel”, американський “Learn to speak English” та власні записи мовлення американців під час стажування в Техаському університеті в Арлінгтоні, США за програмою Фулбрайта (2003-2004). Було вияв-

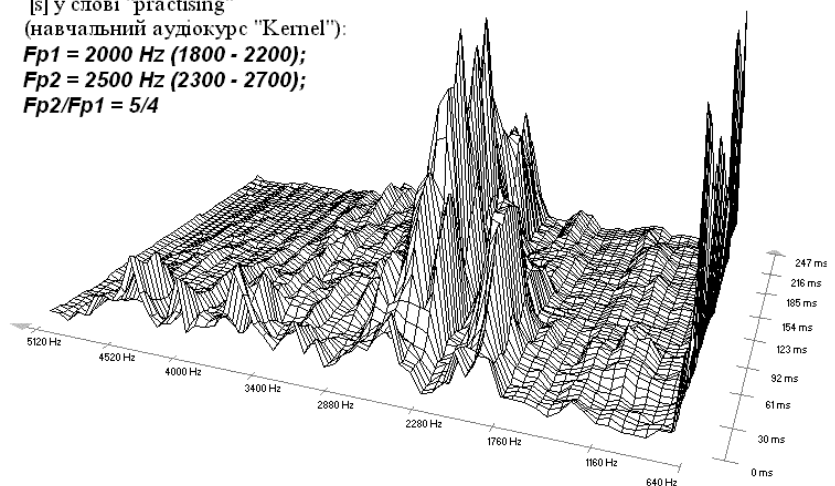
лено, що ця фонема характеризується формантами з відносно стабільними частотами: $Fp1 = 2000 \text{ Hz}$ (1800 – 2200 Hz), $Fp2 = 2500 \text{ Hz}$ (2300 – 2700 Hz) і формант-відношенням $r = Fp2/Fp1 = 5/4$ (велика терція), що відповідає українському [i] (див. мал. 4). Низькочастотна (~300 Hz) побічна форманта, зумовлена РГ, не може бути включена в інваріантні співвідношення.

Звук [i] у спадному тоні (ч):
є дві стабільні форманти
з частотами
 $Fp1 = 2500 \text{ Hz}$
 $Fp2 = 3150 \text{ Hz}$,
 $Fp2/Fp1 = 5/4$



Мал.3

Англійський звук [ɪ] після приголосного
[s] у слові "practising"
(навчальний аудіокурс "Kernel"):
 $Fp1 = 2000 \text{ Hz}$ (1800 - 2200);
 $Fp2 = 2500 \text{ Hz}$ (2300 - 2700);
 $Fp2/Fp1 = 5/4$



Мал.4

У нормальному мовленні спектр [y] визначається насамперед низькочастотним РГ, де $F1_H = 300 - 450 \text{ Hz}$ та $F2_H = 500 - 780 \text{ Hz}$, а відношення частот становить $r_H = 5/3$ (1.70 ± 0.03). При мовленні високим тоном та шепотом, коли виникають обмеження на цей вид резонансу, починають домінувати відносно високочастотні максимуми з тим самим відношенням між їхніми частотами, $r = 5/3$ (представлені в табл. 1). У деяких випадках спостерігалася також додаткова форманта з частотою $F_{add} = 2100 - 2600 \text{ Hz}$, де $Fp1/F_{add} = 5/3$. Під час шепоту виникала також додаткова пара формант, де $Fp1_{ex} = 3000 - 3200 \text{ Hz}$ та $Fp2_{ex} = 5000 - 5200 \text{ Hz}$, $r = 5/3$. Найвірогідніше, всі ці додаткові резонанси зумовлені ТР в одній із порожнин МТ

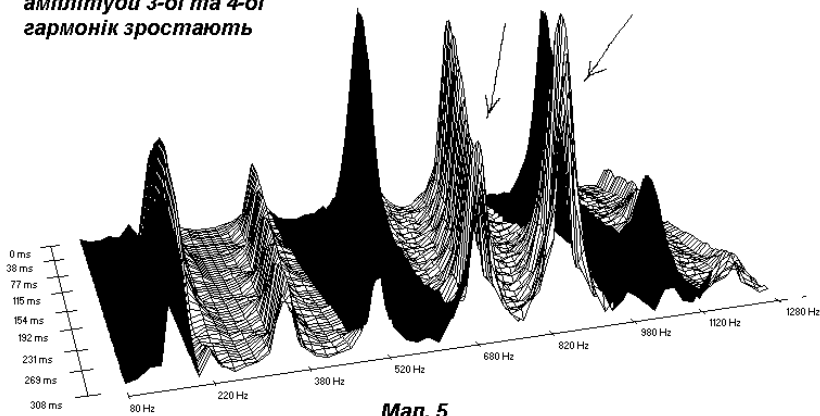
та в звуженні між двома порожнинами.

Коли висота [a] понижується, наступні вищі гармоніки потрапляють в область резонансу, а нижчі загасають (див. мал. 5).

Коли висота тону зростає, пари формант "пересуваються" вище по частотній шкалі. Так, у високому звуку [a] з $F_0 = 800 \text{ Hz}$ (значно вище діапазону Гельмгольца) ми спостерігали основну пару головних формант із частотами $Fp1 = 3000 \text{ Hz}$ (2800–3200 Hz), $Fp2 = 4000 \text{ Hz}$ (3800–4200 Hz) та додаткову, де $Fp1(2) = 6000 \text{ Hz}$ (5700–6300 Hz), $Fp2(2) = 8000 \text{ Hz}$ (7700–8300 Hz). Відношення ФЧ лишилося тим же: $r = 4/3$.

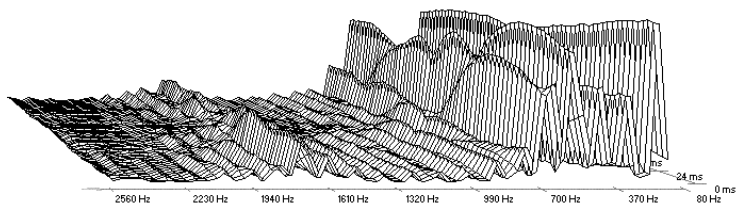
На мал. 6,7 зображені різні варіанти вимови звука [e], де $r = 3$.

Звук [а] у спадному тоні:
при наближенні до
резонансних частот
амплітуди 3-ої та 4-ої
гармонік зростають

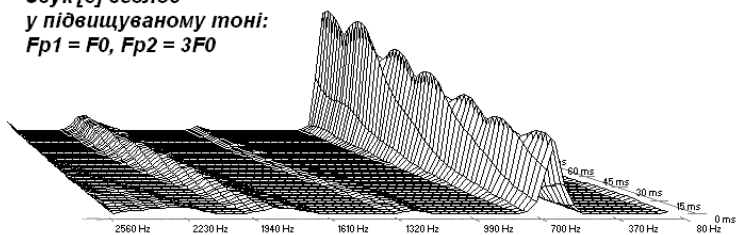


Мал. 5

Звук [е] вголос (ч):
 $F_{p1} = 570 \text{ Hz}$, $F_{p2} = 1710 \text{ Hz}$;
 $F_{p2}/F_{p1} = 3$

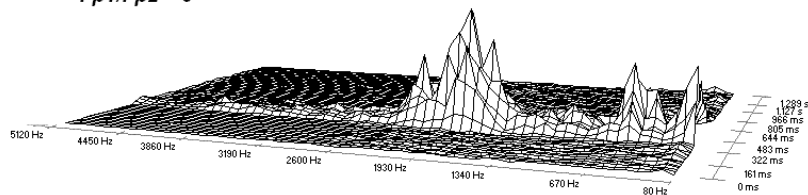


Звук [е] вголос
у підвищеному тоні:
 $F_{p1} = F_0$, $F_{p2} = 3F_0$

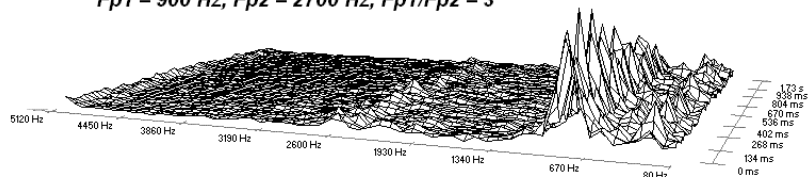


Мал. 6

Звук [е] вголос (ч):
 $F_{p1} = 660 \text{ Hz}$, $F_{p2} = 2000 \text{ Hz}$;
 $F_{p1}/F_{p2} = 3$



Звук [е] пошепки (ч):
 $F_{p1} = 900 \text{ Hz}$, $F_{p2} = 2700 \text{ Hz}$, $F_{p1}/F_{p2} = 3$



Мал. 7

Якщо один звук “змішується” з іншим (як [у] в комбінації [оу] чи у вимові “о”-типу), виникають додаткові максимуми на частотах із відношенням, притаманним “домішаному” звуку. Зокрема, звук [у] з “о-домішкою” набуває додаткових формант на частотах $F1(o) = 2400-3000$ Hz, $F2(o) = 3000-4500$ Hz із відношенням $r(o) = F2(o)/F1(o) = 3/2$.

Відзначимо цікаву особливість звуків мовлення – їх аналогію з музичними інтервалами. Так, звукові [и] відповідає мала (мінорна) терція, [і] – велика (мажорна) терція, [а] – кварта, [о] – квінта, [е] – квінта через октаву. Відтак, наявність звуків [і] та [и] в українській мові – безсумнівний доказ її багатства та розвиненості: адже вона, як цікава мелодія, містить і мажорні, і мінорні лади.

Отже, як ми й очікували, акустичні інваріанти зумовлюються ТР.

Така сталість формант-відношення – вкрай важливий результат, який підтверджує наше теоретичне передбачення про те, що єдиним фонемним інваріантом є відношення частот головних формант.

Проведений аналіз дає змогу створити будь-який звук засобами хвильового синтезу. При цьому штучно створені звуки (синтезовані хвилеформи [а], [і], [и]) мають більш стабільне та чисте звучання, ніж натуральні.

Нами спостерігалися значні відмінності в амплітудах гармонік від мовця до мовця, що добре узгоджується з експериментальними даними [16; 17; 20; 7], які свідчать, що амплітуда гармонік поблизу 2,5 kHz відносно амплітуди першого обертоу може варіювати в інтервалі до 20 dB залежно від мовця. Це дозволяє припустити, що тембр голосу особи проявляється у відношеннях амплітуд гармонік.

Отже, ми дослідили спектральні характеристики українських голосних для різних варіантів вимови: звичайного мовлення, шепоту та змінного тону. На основі інваріантно-акустичного аналізу мовлення були знайдені акустичні інваріанти (формант-інваріанти) цих звуків. Було доведено, що форманти звуків мовлення розпадаються на два класи: головні (постійні, перманентні), які зумовлюють інваріантні акустичні характеристики, та побічні (змінні), які пов’язані з особливостями особи чи вимови. Головні форманти характеризуються інваріантним відношенням їх частот, яке є незмінним для даного звука мовлення. Показано, що в нормальному мовленні найнижчі фонемні частоти можуть бути частотами побічних формант, зумовлених резонансом Гельмгольца. Додаткові резонанси можуть виникати за достатньо високої ЧОТ, коли у резонаторі вміщується кілька вищих обертонів: $n = 1, 2, \dots$, і т.д. У цьому випадку з’являються додаткові форманти на частотах,

пропорційних частотам головних формант. Показано також, що формант-інваріанти українського звука [і] та англійського [I] однакові.

Найважливіший висновок цієї праці такий: єдиним фонемним інваріантом є відношення між частотами головних формант, а не їх абсолютні значення. Одержані результати можуть стати в нагоді спеціалістам у галузі експериментальної фонетики та моделювання мовлення.

Ця робота частково підтримана грантом №68427736 програми Фулбрайта на 2003-2004 академічний рік.

Література

1. Трубецкой Н.С. Основы фонологии. – М.: Изд-во иностр. лит. – 1960. – С. 7-22.
2. Вакуленко Максим. Правописні аспекти науки термінології. – Вісник Книжкової палати, 1998, №11. – С. 15-17.
3. Studies in Communicative Phonetics and Foreign Language Teaching Methodology / M.P.Dvorzhetska, A.A. Kalita, Eds. Lenvit. – Kyjiv, 1997.
4. Vakulenko M.O. Transliteration Through a Slavonic Latin Alphabet: Saving Information and Expenses. – Вісник Київського лінгвістичного університету. – Серія Філологія. – Т.2, №1 (1999). – С. 85-94.
5. Stevens K.N. Acoustic Phonetics. – MIT Press, 1998. – 607 p.
6. Flanagan J.L. Speech analysis, synthesis, and perception. – Berlin: Springer-Verlag, 1972.
7. Klatt D.H., Klatt L.C. Analysis, synthesis, and perception of voice quality variations among female and male talkers // J. of the Acoustical Soc. of America. – 1990. - № 87. – P. 820-857.
8. Traunmueller H. Perceptual dimensions of openness in vowels // J. of the Acoustical Soc. of America. – 1981. - № 69. – P. 1465-1475.
9. Hoemeke K.A., Diehl R.L. Perception of vowel height: The role of F1-F0 distance // J. of the Acoustical Soc. of America. – 1994. - № 96. – P. 661-674.
10. Carlson R., Granstroem B., Fant G. Some studies concerning perception of isolated vowels // Speech Transmission Laboratory Quarterly Progress and Status Report 2-3, Royal Institute of Technology. – Stockholm, Sweden, 1970. – P. 19-35.
11. Menard L., Schwartz J.-L., Boë L.-J., Aubin J. Articulatory-acoustic relationships during vocal tract growth for French vowels: Analysis of real data and simulations with an articulatory model // J. of Phonetics. – 2007. - №35. – P. 1-19.
12. Nearey T.M., Assmann P.F. Information conveyed by f0 for vowel identification // J. of the Acoustical Soc. of America. – 2006. - № 119. – P. 33-39.
13. Методы автоматического распознавания речи: в 2-х кн. Пер. с англ. / под ред. У.Ли. – М.:

- Мир, 1983. – Книга 1. – 328 с., ил.
14. Винцюк Т.К. Анализ, распознавание и интерпретация речевых сигналов / Т.К. Винцюк. – К., 1987. – 262 с.
 15. Fant G. Acoustic theory of speech production. – The Hague, Netherlands, 1960.
 16. Stevens K.N., Hanson H.M. Classification of glottal vibration from acoustic measurements // Vocal fold physiology: Voice quantity control / O.Fujimura, M.Hirano, Eds. – San Diego: Singular, 1995. – P. 147-170.
 17. Hanson H.M. Glottal characteristics of female speakers. PhD dissertation. – Harvard University, Cambridge MA, 1995.
 18. Fujimura O., Lindqvist J. Sweep-tone measurements of vocal-tract characteristics // J. of the Acoustical Soc. of America. – 1971. - № 49. – P. 541-558.
 19. Fant G., Nord L., Branderud P. A note on the vocal tract wall impedance // Speech Transmission Laboratory Quarterly Progress and Status Report 4. - Royal Institute of Technology, Stockholm, Sweden, 1976. – P. 13-27.
 20. Holmberg E.B., Hillman R.E., Perkell J.S. Glottal airflow and transglottal air pressure measurements for male and female speakers in soft, normal and loud voice // J. of the Acoustical Soc. of America. – 1988. - № 84. – P. 511-529.
 21. Zhang Zh., Neubauer J., Berry D.A. The influence of subglottal acoustics on laboratory models of phonation // J. of the Acoustical Soc. of America. – 2006. - № 120. – P. 1558-1569.
 22. Тоцька Н.І. Сучасна українська літературна мова: фонетика, орфоепія, графіка, орфографія. – Київ: Вищ. шк., 1981. – 183 с.
 23. Вакуленко М.О. Аналіз та синтез звукових спектрів людського мовлення. – Пульсар. – 1999. - № 6-7. – С. 20-23.
 24. Вакуленко М.О. Акустичні характеристики та інваріанти звуків української мови. – Науковий вісник кафедри ЮНЕСКО КДЛУ. – Філологія, педагогіка, психологія. – Вип. 1. Київ, 2000. – С. 62–66.
 25. Dubeda T., Keller E. Microprosodic aspects of vowel dynamics – an acoustic study of French, English and Czech // J. of Phonetics. – 2005. - № 33. – P. 447-464.
 26. Фейнмановские лекции по физике. Тт. 3, 4: Излучение, волны, кванты. Кинетика, теплота, звук // Р.Фейнман, Р.Лейтон, М.Сэндс / пер. с англ. А.В.Ефремова, Г.И.Копылова, Ю.А.Симонова, О.А.Хрусталева / под ред. Я.А.Смординского. – М.: Мир, 1976. – 496 с.

On the basis of invariant acoustic speech analysis, the permanent spectral characteristics of Ukrainian vowels are obtained for different pronunciation modes: normal speech, whisper and varying tone. It is shown that the acoustic invariants of speech sounds are the relations between frequencies of principal formants, not their absolute values. It is demonstrated that the acoustic invariants of Ukrainian sound [i] are identical to those of English sound [I].

Key words: acoustic phonetics, speech recognition, speech sounds, acoustic parameters, acoustic invariants.

На основе инвариантно-акустического анализа речи получены постоянные спектральные характеристики украинских гласных для различных способов произношения: обычной речи, шёпота и переменного тона. Показано, что акустическими инвариантами звуков речи являются отношения между частотами главных формант, но не их абсолютные значения. Продемонстрировано, что акустические инварианты украинского звука [i] и английского [I] одинаковы.

Ключевые слова: акустическая фонетика, распознавание речи, звуки речи, акустические параметры, акустические инварианты.