

УДК 681.84.087.4

## ЗБЕРЕЖЕННЯ ПРИРОДНИХ ТЕМПОРАЛЬНИХ ЗАКОНОМІРНОСТЕЙ У ПРОЦЕСАХ СПОВІЛЬНЕНОГО ВІДТВОРЕННЯ МОВНОЇ ІНФОРМАЦІЇ

© Шпак З., Рашкевич Ю., 2003

*Наведено результати досліджень часових параметрів одиниць мовного потоку для різних темпів мовлення. Проаналізовано закономірності темпоральних змін звуків української мови та їх складових частин. Виділено групи мовних елементів, які характеризуються спільністю темпоральних перетворень, викликаних зміною швидкості мовлення. Побудовано графіки, що відображають залежність відносної зміни тривалості елементів виділених груп від загального коефіцієнта зміни темпу.*

*The characteristics of the speech units duration have been specified for different pronunciation types. Time-scale transformation regularity of the Ukrainian voice phonemes sounding at different speech rates has been investigated. Six groups of oral speech elements are determined: stressed vowels, unstressed vowels, voiced consonants, unvoiced consonants, plosive stops and word-spacing pauses. The elements of specified groups are characterized by similar temporal transformations. Non-linear analytic functions presenting the relative temporal changes of sound duration have been plotted for every group.*

### 1. Вступ

Керування темпом надходження голосових повідомлень дозволяє істотно підвищити ефективність усномовної комунікації в людино-машинних системах. Для забезпечення високої якості звучання мови, відтвореної зі зміненою швидкістю, процеси часового перетворення повинні базуватись на природних темпоральних властивостях мовних сигналів.

Практично всі напрацювання зарубіжних спеціалістів, присвячені проблемі часового масштабування мовних сигналів, передбачають рівномірну модифікацію тривалості голосових записів без урахування особливостей темпоральних змін окремих звуків [1].

У роботі [2] запропоновано новий підхід, який базується на часовій трансформації мовного потоку, адаптованій до його звукового складу, і на основі цього підходу розв'язано задачу максимального скорочення часу відтворення мовних записів. Проте задача сповільнення швидкості надходження мовних повідомлень вимагає самостійного розгляду, оскільки, по-перше, природні процеси зменшення темпу мовлення характеризуються специфічними закономірностями перетворень структур звуків, по-друге, першочерговим завданням сповільнення є забезпечення високої розбірливості та натураль-

ності звучання перетвореного мовного потоку, а також збереження просодичних особливостей вимови.

Тому актуальним питанням розроблення високоякісних методів керування швидкістю відтворення мовної інформації є дослідження природних темпоральних закономірностей, притаманних голосовим сигналам, зокрема для української мови.

## 2. Визначення часових характеристик різних темпів мовлення

Для визначення загальних темпоральних параметрів мовлення використано набір текстів, які відображають сучасні тенденції у структурі української розмовної, ділової та наукової мови. Кожен з текстів (загальний обсяг мовного матеріалу становив 167 речень, відповідно: 2138 слів, 5331 складів і 12315 фонем) промовлявся групою з семи

Таблиця 1

Часові характеристики основних стилів мовлення

Параметр		Темп мовлення								
		швидкий			звичайний			повільний		
		M1	M2	сер.	M1	M2	сер.	M1	M2	сер.
Середня тривалість	речення (с)	8,16	5,29	5,79	12,43	5,10	7,57	20,34	9,69	12,08
	синтагми (с)	2,37	1,47	1,68	3,53	2,34	2,47	5,64	2,44	3,45
	слова (мс)	552	341	415	734	441	540	1106	556	667
	складу (мс)	221	131	159	302	173	206	419	213	267
	звуку (мс)	94	56	68	129	74	88	179	91	114
Швидкість мовлення	слів/хвилину	96,5	135,8	130,1	64,6	119,7	89,1	39,2	86,1	62,3
	складів/секунду	4,1	6,7	5,3	2,6	4,7	3,7	1,6	3,4	2,5
	звуків/секунду	9,2	16,3	13,2	6,3	11,5	9,0	3,7	8,2	6,2
Коефіцієнт прискорення/сповільнення		1,51	1,43	1,46	1,0	1,0	1,0	1,64	1,39	1,43

мовців жіночої та чоловічої статі в чотирьох темпах: швидкому, звичайному розмовному, повільному та протяжному. У табл. 1 наведено темпоральні дані двох мовців (M1 і M2), які істотно відрізняються стилями мовлення, та усереднені значення для всієї групи мовців.

Аналіз отриманих результатів дозволив зробити такі основні узагальнення:

- найбільш стабільною і незалежною від тексту характеристикою швидкості мовлення є параметр "складів за секунду" – його доцільно застосовувати як базову темпоральну характеристику;
- варіативність темпоральних показників мовців є достатньо широкою, особливо у випадках швидкої та повільної вимови. Так, M1 володіє достатньо широким діапазоном зміни темпу і помірною швидкістю мовлення, а M2 може бути віднесений до групи швидких мовців – властивий для нього звичайний темп є близьким до швидкого мовлення M1;
- незважаючи на відмінності в абсолютних параметрах швидкості мовлення, всім мовцям властиві достатньо близькі значення відносної зміни темпу. Середні за всією групою коефіцієнти максимального прискорення та сповільнення становили відповідно 1,46 та 1,48 (коефіцієнти сповільнення характеризувались значно ширшим розкидом значень, ніж коефіцієнти прискорення);
- часові параметри мовця не залежать від його статі, а визначаються властивим даній людині стилем мовлення. Усім мовцям, які не мали спеціальної дик-

торської практики, важко було довготривало відхилятися від звичайної швидкості вимови – чим довшим був текст, тим меншим ставав загальний коефіцієнт зміни темпу.

### 3. Аналіз темпоральних перетворень пауз і звуків

#### 3.1. Вплив темпу мовлення на характеристики пауз

Зміна темпу мовлення пов'язана з перерозподілом структури мовного потоку (рис. 1). Перехід від звичайної мови до швидкої чи повільної насамперед відображається в зміні кількості та тривалості пауз: кількість пауз зменшується/зростає в 1,5 разів, а їх

Темпоральні характеристики розділчих пауз

Таблиця 2

Характеристика	Темп мовлення			
	швидкий	звичайний	повільний	протяжний
Тривалість пауз у 1с мовного потоку (мс)	95	198	293	214
Кількість пауз на слово	0,41	0,56	0,81	0,97
Середня тривалість паузи (мс)	141,6	238,1	383,4	405,7
СКВ тривалості пауз (мс)	129,1	188,7	215,3	273,7
Середня тривалість звукової ділянки між паузами (мс)	1012	664	905	1420
Відносне збільшення загальної тривалості пауз	1,0	2,47	5,73	6,41
Відносне збільшення середньої тривалості паузи	1,0	1,67	2,72	2,84

сумарна тривалість змінюється відповідно приблизно в 2,3 раза (табл.2). Водночас загальна тривалість звукової частини мовного потоку скорочується/збільшується тільки в 1,3 раза. Найбільший відсоток пауз у повільній мові – близько 30%. Зазначено, що при подальшому сповільненні темпу дещо зростає кількість пауз, але їх середня тривалість майже не змінюється. Як результат – у протяжній мові частка пауз зменшується до 20%.



Рис. 1. Відсоток пауз і звуків у загальній тривалості мови

значення тривалості звукової ділянки між паузами, яке для експериментальних даних було близьким до 0,96 с, зазнає найменших темпоральних змін. Тільки в протяжному, близькому до співочого мовленні цей параметр зростає в 1,5 раза.

#### 3.2. Особливості темпоральних перетворень різних звуків

Для аналізу темпоральних перетворень звуків виконано аудіовізуальний поділ мовних сигналів кожного із записів. Границі звуків контролювались та уточнювались за динамічними спектрограмами, що дозволило виділити складові частини звуків: початкові та кінцеві переходи і серединні квазістаціонарні ділянки.

Основна різниця зазначена в темпоральній поведінці голосних і приголосних (передусім шумних) звуків. Для всіх мовців сповільнення темпу проявлялось у зростанні частки голосних звуків: співвідношення голосні – приголосні для швидкого темпу

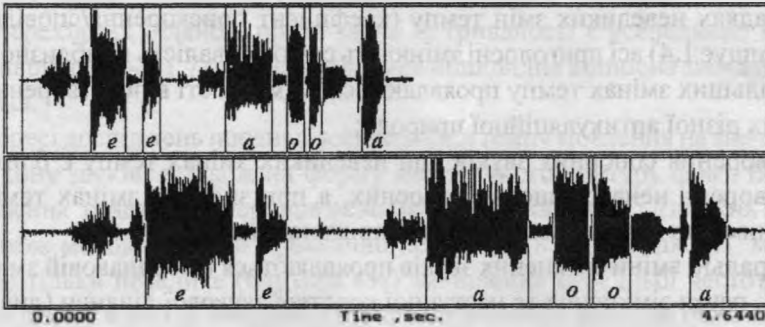


Рис.2. Мовні сигнали словосполучення щебет жайворонка, вимовленого в звичайному та протяжному темпах

становило 0,42:0,58; для звичайного – 0,45:0,55; для повільного – 0,51:0,49; а для протяжного – 0,59:0,41.

Перехід до швидкого чи повільного мовлення найбільше проявляється у зміні тривалості голосних звуків, передусім наголошених (рис.2), проте в протяжній вимові, для якої голосні звуки відіграють домінуючу роль, спостерігається певне зменшення інтонаційного співвідношення між наголошеними і ненаголошеними звуками. Місце

Таблиця 3  
Характеристики темпоральної тривалості звуків базових класів

Класи звуків		Темп мовлення												
		швидкий			звичайний			повільний			протяжний			
		ПВ, %	СТ, мс	ВЗТ	ПВ, %	СТ, мс	ВЗТ	ПВ, %	СТ, мс	ВЗТ	ПВ, %	СТ, мс	ВЗТ	
Голосні	ненаголошені	22,6	59,5	1,32	22,8	78,8	1,0	25,0	105,8	1,34	29,6	239,5	3,04	
	наголошені	19,5	91,1	1,42	21,9	129,6	1,0	25,2	181,5	1,40	29,8	408,5	3,15	
	сонорні	20,6	49,6	1,30	19,9	64,7	1,0	17,5	84,6	1,31	15,1	119,9	1,85	
Приголосні	шлівні	дзвінки	4,8	61,7	1,27	4,6	78,6	1,0	4,1	98,9	1,26	3,4	136,8	1,74
		глухі	8,9	99,4	1,26	8,7	125,6	1,0	7,7	146,5	1,17	5,3	168,1	1,34
	зімкнені	дзвінки	4,8	36,5	1,32	4,7	48,0	1,0	4,0	60,9	1,27	3,2	79,5	1,66
		глухі	2,6	24,6	1,25	2,5	30,7	1,0	2,1	36,6	1,19	1,2	42,3	1,38
	змикання	10,2	71,0	1,27	10,6	90,1	1,0	10,8	109,1	1,21	8,3	146,4	1,62	
	африкати	2,8	77,1	1,28	2,6	98,4	1,0	2,4	110,8	1,13	1,7	127,6	1,30	
	подовжені	1,7	108,7	1,31	1,7	142,4	1,0	1,6	178,3	1,25	1,3	216,8	1,52	

Примітка: ПВ - процентний вміст; СТ - середня тривалість; ВЗТ - відносна зміна (збільшення/зменшення) тривалості

голосного звуку в слові та просодика цього слова мають певний вплив на абсолютну тривалість звуку, проте значення відносних змін тривалостей, викликаних зміною темпу мовлення, для всіх голосних звуків є загалом достатньо близькими.

Відмінність у темпоральних перетвореннях голосних і приголосних звуків набуває принципового характеру у випадках значного сповільнення темпу (табл. 3). Так, перехід від швидкого до протяжного мовлення пов'язаний із загальним збільшенням тривалості голосних у 4,4 раза, а приголосних – тільки у 2,2 раза. Водночас темпоральна

поведінка різних приголосних звуків є теж достатньо неоднорідною. Проведено аналіз темпоральних змін для кожного звуку зокрема та виконано узагальнення для основних класифікаційних груп приголосних [2]. У межах кожної групи враховувався вплив факторів пом'якшення, дзвінкості, назалізації та інших. Зроблено такі основні висновки:

- у випадках невеликих змін темпу (коефіцієнт прискорення/сповільнення не перевищує 1,4) всі приголосні змінюють свою тривалість приблизно однаково, при більших змінах темпу проявляються відмінності в перетвореннях приголосних різної артикуляційної природи;
- перетворення сонорних звуків при невеликих змінах темпу є близькими до перетворень ненаголошених голосних, а при значних змінах темпу вони є ближчими до перетворень дзвінких приголосних;
- темпоральні зміни зімкнених звуків проявляються в неоднаковій зміні їх складових: паузи зімкнення та наступної короткої звукової ділянки (див. табл. 3);
- як для сонорних, так і для щільних та зімкнених звуків не встановлено істотних відмінностей у темпоральних змінах тривалостей м'яких та відповідних твердих звуків, за винятком випадків, коли м'які приголосні виступали як подовжені;
- приголосні з тональною складовою (дзвінкі звуки) дещо відрізняються за параметром абсолютної тривалості, а також за значеннями відносних темпоральних змін від відповідних глухих приголосних (табл. 3), особливо у випадках протяжного мовлення;
- збільшення тривалості початково довгих звуків (зокрема подовжених приголосних) і скорочення початково коротких (зокрема вибухових) відбувається меншою мірою, ніж для інших звуків даної групи.

### 3.3. Виділення основних темпоральних груп

Узагальнення результатів показало, що незважаючи на розкид значень абсолютної тривалості окремих звуків і певну варіативність їх темпоральних змін, загалом звуки

зі спільними акустико-артикуляційними механізмами формування характеризуються подібністю значень відносних змін тривалостей (табл. 3). З усієї множини звуків української мови виділено чотири основні темпоральні групи: наголошені голосні, ненаголошені голосні, тональні (сонорні та дзвінкі) приголосні, шумні (глухі) приголосні. Дві окремі групи складають паузи зімкнення та роздільчі (міжслівні) паузи, темпоральні модифікації яких є повністю відмінними. На рис.3 наведено графіки, що характеризують середні значення зміни тривалості мовних елементів основних темпоральних груп відносно значень, властивих для цих елементів у швидкому темпі:  $k$  – загальний коефіцієнт сповільнення вимови,  $s$  – відносна зміна тривалості звуків/пауз даної темпоральної групи.

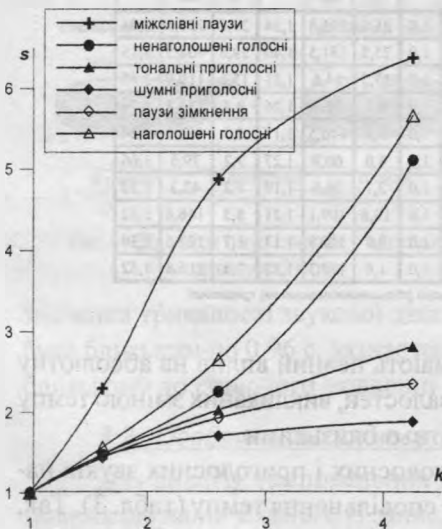


Рис.3. Графіки відносних змін тривалості елементів базових темпоральних груп

Зміна швидкості мовлення викликає неоднакові перетворення стаціонарної і перехідних ділянок звуків. Якщо зміни тривалості стаціонарів є достатньо регулярними, то трансформації перехідних ділянок характеризуються великою варіативністю навіть для однакових звуків. У більшості випадків змінюються тривалості як початкової, так і кінцевої перехідних ділянок, проте зміна їх тривалості є усереднено в 1,4 а в разі значних сповільнень – в 1,7 раза меншою, ніж відповідна відносна зміна стаціонарних частин звуків.

У процесі досліджень проаналізовано вплив темпу мовлення на значення частоти основного тону звуків, а також на форму мелодичного контуру фраз. Виявлено, що середні значення тривалості періодів основного тону голосних і дзвінких приголосних звуків, а також мелодика мови практично не залежать від швидкості мовлення [4]. Зафіксовано тільки невелике (близько 4%) зменшення середньої частоти основного тону у протяжному темпі та модуляції значень тривалості періодів основного тону для окремих довгих голосних. Темпорально стабільними виявилися також розподіли формантних частот і спектральні структури шумних звуків.

#### 4. Висновки

Природна зміна швидкості мовлення пов'язана зі зміною тривалості всіх елементів мовного потоку. У разі невеликих змін темпу основних трансформацій зазнають паузи, а модифікації всіх звуків є достатньо подібними і можуть бути описані лінійною залежністю від коефіцієнта зміни темпу. У випадках сповільнення/прискорення більше, ніж у 1,4 рази, проявляються розбіжності у темпоральних перетвореннях звуків з різними механізмами формування, передусім між голосними і приголосними звуками. Ці відмінності стають принциповими для протяжної мови.

За близькістю значень відносної зміни тривалості і подібністю перетворення структур сигналів, викликаних зміною швидкості мовлення, всі звуки української мови можна об'єднати в чотири основні темпоральні групи: наголошені голосні, ненаголошені голосні, тональні приголосні та шумні приголосні; ще дві окремі групи формують паузи: міжслівні та вибухові. Встановлено, що залежність значень зміни тривалості елементів кожної з цих груп від величини загального коефіцієнта сповільнення має нелінійний характер, що особливо проявляється для голосних звуків у випадках великих змін темпу мовлення. Властивість обмежених часових перетворень надмірно довгих чи коротких звуків можна враховувати через додаткове введення функцій нормування тривалості [5].

Зроблено загальний висновок, що реалізацію зміни швидкості відтворення мовних записів за умови натуральності звучання голосового повідомлення необхідно здійснювати шляхом диференційованої зміни тривалості мовних елементів на основі закономірностей, властивих їх темпоральним групам, зберігаючи загальну спектральну структуру звукових сигналів, передусім тональних.

1. Laroche J., Dolson M. Improved Phase Vocoder. Time-Scale Modification of Audio // IEEE Transactions on Speech and Audio Processing. – 1999. – Vol.7, №3. – P. 323 – 322.
2. Рашкевич Ю.М. Перетворення часового масштабу мовних сигналів. – Львів: Академічний експрес, 1997. – 140 с.
3. Тоцька Н.І. Сучасна українська літературна мова. – К.: Вища школа, 1981. – 183 с.