Софийски университет "Св. Климент Охридски"
Факултет по математика и информатика
Катедра "Математическа логика и приложенията ѝ"

Дипломна работа

Разпознаване на емоции в сигнали от реч и ЕЕГ

Диана Генева

Специалност "Компютърна лингвистика"

Факултетен номер: 25742
Email: dageneva@qtrp.org

Научен ръководител: Доц. Стоян Михов

2019

Съдържание

Нулева зона

Много хора си мислят, че в началото бе Словото. Всъщност,

> В началото бе създадена Вселената. Този факт разгневи силно много хора и сега се шири мнението, че това е била погрешна стъпка.[1]

След това дойде словото.

Словото, и по-точно говоримата реч, все още е най-ефективният метод за предаване на информация между хора. Това е така, защото речта е изключително удобна за тази цел. От една страна, защото е естествена, в смисъл на "хората са анатомично предразположени да произвеждат реч". От друга, може да се комуникира, без да има нужда от директен зрителен контакт, на малки или сравнително големи разстояния. Ако ситуацията позволява, може да се кодира и допълнителна информация. Обикновено когато се говори за реч, се различават два компонента. Единият, **очевидният**, е словесният, който предава експлицитно съобщение. Вторият е компонентът на прозодията. Прозодията се отнася до елементи на речта като интонация, тон, ритъм и ударение. Чрез нея се кодира и допълнителна информация, която не може да се предаде чрез граматика или писмен вид. Прозодията може да определя дали изречението е въпросително или е повелително, дали е иронично или саркастично. Най-вече, по този начин се кодира имплицитна информация за говорещия и неговото емоционално състояние. Добавянето на този канал към комуникацията я прави значително по-богата и много по-ефективна. В много ситуации, без информация за прозодията на някакво изказване, дори не може да се предаде съобщението му (Фигура 0.0.1). Поради тази причина, всяко приложение, използващо човешка реч, е желателно да се възползва и от информацията, кодирана в прозодията. Хората са много чувствителни към нея. Показано е в [Wee+00], че има голяма разлика във възприетата важност на съобщението в зависимост от интонацията, както и че например пилотите предпочитат човешки пред компютърен глас при предаване на важни съобщения[2] [BSS00]. Тъй като голяма част от прозодията кодира информация за емоционалното състояние на говорещия, разпознаването на емоцията в сигнали е интересен въпрос.

Съществуват много изследвания в областта на разпознаването на емоции от мозъчни вълни, както и съществуват много мулти-модални системи, които съчетават вторични изрази като реч, визуални данни, термални данни и

[1] Ресторант „На края на Вселената" от Дъглас Адамс

[2] Интересен факт: жаргонният израз за женски глас, използван за предупредително съобщения в самолети, е "Bitching Betty"

други. Наличието на подръчен електроенцефалограф, лесното създаване на аудио данни и естественото любопитство вдъхновиха опита за съчетаване на един първичен канал - сигнал от електроенцефалограф - и един вторичен - сигнал от реч. Въпросът, който се разглежда, е до какво ще доведе съчетаването на тези два сигнала при разпознаването на емоции в тях. Текстът на дипломната работа е разделен на значещите части на заглавието ѝ както следва:

- В Глава 1 е описано какво все пак имаме предвид под емоция и какво точно целим да разпознаваме
- В Глава 2 е описан физичният процес на производство на реч и характеристичните вектори, които ще извличаме от този сигнал, както и класификационни резултати само от този сигнал
- В Глава 3 са описани (с много по-малко детайли) спецификите на извличане на характеристични вектори от ЕЕГ сигнал, както и класификационни резултати само от него
- В Глава 4 се описват методи за съчетаване на двата сигнала и резултатът от това
- Накрая заключението е изложено в Големият портрет

Фигура 0.0.1: Пример за многозначност на изказване, която не може да бъде разрешена, без да се предаде допълнителна информация.

Попълването на вица е оставено за упражнение на читателя

Глава 1

Разум и чувства

За да кажем как ще разпознаваме емоции в сигнали, трябва да изберем две неща. Първо, какво ще наричаме емоция и второ - кои емоции ще класифицираме.

Най-точното описание, което може да се даде за емоция, е може би цитатът на Потър Стюърд (член на Върховния съд на САЩ): "Когато го видя, ще го разпозная"[1]. Изкушаващо е да се използват за случая думите му: "Няма днес да се опитвам точно да дефинирам материята, попадаща под това кратко описание; възможно е никога да не мога да дам разбираемо описание. Но едно нещо знам, когато го видя, ще го разпозная.". За да можем да продължим напред в текста на дипломната работа обаче, ще трябва да работим с малко по-конкретни дефиниции.

Един от най-ранните опити да се дефинира емоция идва от Чарлз Дарвин и книгата му "За изразяването на емоциите при човека и животните" от 1872 г. Той използва емоцията, за да потвърди еволюционната си теория, тъй като според неговите наблюдения изразяването на емоциите е сходно между животните и хората и се обяснява еволюционно. В книгата си Дарвин представя три принципа, които описват изразяването на емоциите:

1. "Принцип на полезните навици"

   Такива са смръщване на вежди, при което в очите влиза по-малко светлина, или вдигане на вежди, което увеличава полето на зрение. Когато човек е учуден, той вдига вежди, за да "види по-ясно ситуацията". Такива навици, според Дарвин, се предават наследствено заради полезния си характер.

2. "Принцип на противоположностите"

   Това са навици, които нямат практическа полза, но представляват противоположност на някакъв друг естествен навик. Пример за това е свиването на рамене, което е израз противоположен на уверено или агресивно изразяване.

3. "Принцип на нервните сигнали"

[1] I know it when I see it

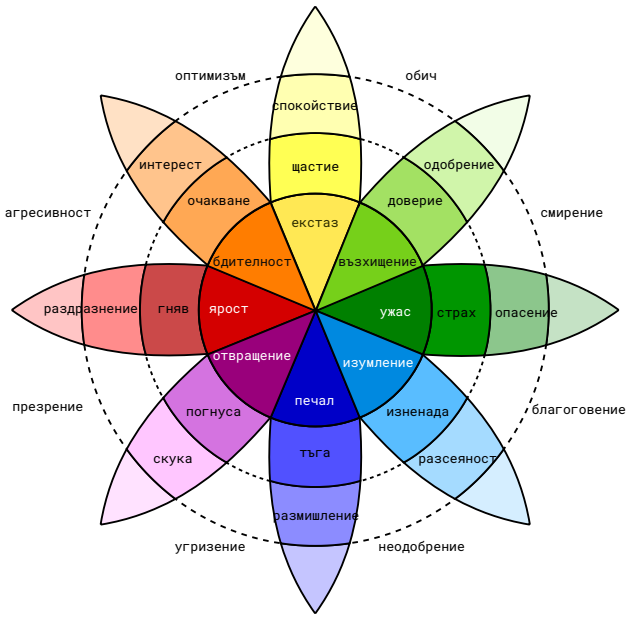
Когато се говори за Дарвиновата теория, най-често се има предвид този принцип. Според него, някои навици са породени от сигнал от нервната система. Такива са например изявите на гняв и страх.

Тъй като тези изрази на емоция се предават еволюционно, то те се наблюдават у всички животни (в частност и хора).

Тази представа за емоциите е продължена от Робърт Плутчик (1927 - 2006). Базирайки се на еволюционната теория, той се опитва да опише понятието "емоция". Наблюденията му в [Plu01] са следните:

Еволюционният произход на някои емоции е по-явен от при други. Например изразът на страх при животните и хората е много сходен и цели едно и също - да премахне причинителя. Изразът на любов обикновено е с репродукционни цели. Някои емоции обаче са по-сложни и първоизточникът им се описва по-трудно, затова търсим начин да генерализираме понятието. Емоцията е сложна верига от събития, която започва с някакъв стимул. В следствие настъпва фаза на "изпитване на емоция" и фаза на физиологични промени (тук възниква научният спор за кокошката и яйцето - дали чувството е първо или физиологичните промени). Те предизвикват целенасочено държание, което цели да премахне дразнението на стимула и да върне състоянието на еквилибриум.

Плутчик прави психо-еволюционна класификация на емоциите, като избира осем главни емоции - две по две противоположни. Всяка от тези базови емоции е свързана с поведение, важно за оцеляването. Например страх и поведението "бий се или бягай", свързано с освобождаването на хормони, подготвящи организма за бягство или битка. Всяка от не-базовите емоции се получава като комбинация на базовите. Плутчик сравнява това с комбинирането на цветове, което може да се види и на известното му "Колело на емоциите" (1980), показано на Фигура 1.0.1.



Фигура 1.0.1: Колелото на емоциите на Плутчик

Базовите емоции са: страх, изненада, тъга, погнуса, гняв, очакване, щастие, доверие. По "листенцата" на всяка от тях се намират по-силният и съответно по-слабият вариант. За съжаление, колкото повече класове искаме да класифицираме, толкова повече данни са нужни, което прави използването на класификацията на Плутчик трудно за практически цели. Друг проблем при голямото разнообразие на Плутчиковото колело е, че различните емоции имат различна продължителност на проявлението, като например тъгата може да продължава с дни, докато изпитването на погнуса е много по-моментно. Това прави съставянето на база данни за емоционално разпознаване трудно.

След като вече започнахме с цитат, време е и за другото клише, а именно етимология. Думата емоция произлиза от френското émouvoir (вълнувам, възбуждам интерес) и по-назад латинското ēmoveō (местя навън/извън пътя). Затова е и естествена класификацията на база как "вълнува" съответната емоция. Известна такава класификация е VAD моделът, разработен от Алберт Мейерабиан и Джеймс Ръсел [Meh74], в който всяка емоция се описва в тримерно пространство с оси "валентност", "ниво на възбуда", "доминантност"[2]. Оста за валентност определя дали емоцията е приятна или не, оста за нивото на възбуда описва колко енергия е нужна за изразяването на емоцията, а доминантност - колко доминиращ или съответно покòрен се чувства човек под влиянието на емоцията. По-често се използва опростен модел, наречен валентност-активация, който използва само две от осите (нивото на възбуда наричаме активация). Изразяването на някои емоции в този модел е показано на Фигура 1.0.2

За да изберем кои емоции да разпознаваме, нека разгледаме какви са силните и слабите страни на двете изследвани области. Според [EKK], в речта лесно се разпознават характеристики, свързани с активацията, тъй като нивото на активацията се отразява на енергията на говора. Тоест емоции като щастие и гняв (които са с висока активация) карат хората да говорят много по-силно и възбудено, докато емоции като тъга и умора (с ниска активация) са свързани с много по-пасивно поведение като цяло. Дадените примери за разпознаване на емоциите в мозъка чрез ЕЕГ сигнал в [AF17] показват, че по-лесно се разпознава валентност, макар че може да бъде извлечена информация и за активацията.

[2]Valence, Arousal, Dominance

Фигура 1.0.2: Модел "валентност-активация"

При избора на емоции, които да изследваме, трябва да изберем такива, за класификацията на които ще помогнат и двата сигнала - тоест да има примери в различните квадранти (ако е възможно) на VAD-модела. Тъй като основните емоции на Плутчик са твърде много, трябва да изберем тяхно подмножество. В случая ще разглеждаме четири емоции - три от които са част от базовите, а четвъртата емоция е "неутрална емоция". Изборът е такъв, че хем да нямаме прекалено много класове, хем тези емоции да могат да се предизвикват лесно в опитна среда и да могат да се описват вербално. Финално избраните емоции, които ще се опитваме да разпознаваме в сигнали от реч и ЕЕГ, са следните:

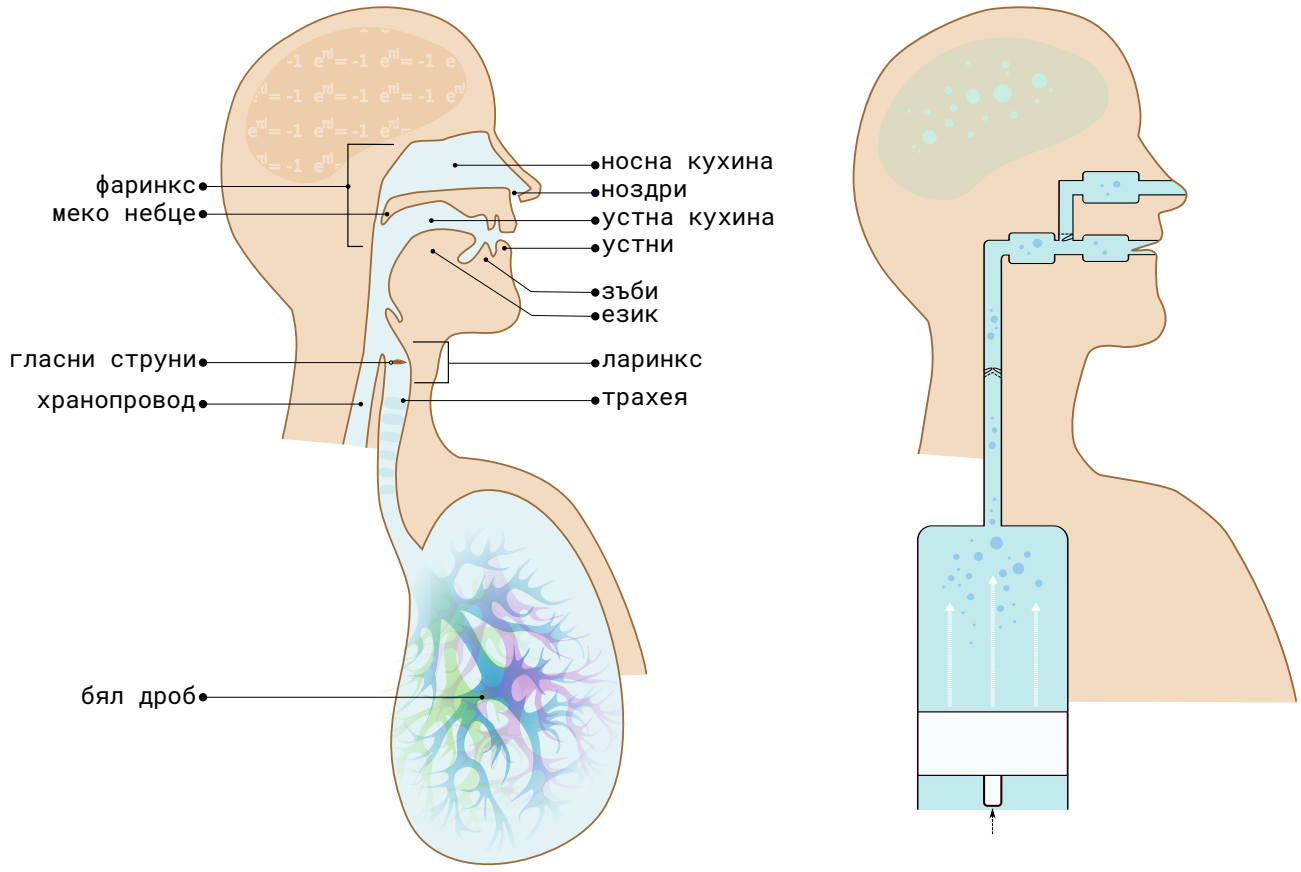
1. **Гняв** - висока активация, отрицателна валентност
2. **Щастие** - висока активация, положителна валентност
3. **Неутрална емоция** - неутрална активация, неутрална валентност
4. **Тъга** - ниска активация, отрицателна валентност

След този бегъл опит да дефинираме понятието емоция и да изберем кои класове емоции ще изследваме, трябва да разгледаме свойствата на двата входни сигнала. Нека започнем от сигналите от реч.

Глава 2

Сигнал от реч

2.1 Физика на тъгата



(а) Физическо представяне на вокалния тракт

(б) Опростено представяне на вокалния тракт

Фигура 2.1.1: Система за производство на реч

Вокален тракт е общото название на кухините над ларинкса (гръкляна), през които минава въздухът при произвеждане на реч. При хората той се състои от ларингеална кухина (съдържаща ларинкса и гласните струни), фаринкс, устна кухина и носна кухина, както може да се види на Фигура 2.1.1а. Вокалният тракт е отговорен за генериране на различни звуци, като текущата конфигурация на отделните му компоненти определя какъв ще бъде самият

звук. Според [KNS09], освен от вида на този звук, конфигурацията на вокалния тракт зависи и от емоцията, която изпитва говорещият. Смята се, че емоционалното състояние е пряко свързано с определени промени в организма, например ускорено дишане или мускулно напрежение, а тези промени се отразяват върху произведената реч. Често дори ефектите от тези промени са станали нарицателно за самата емоция. Из българската литература се срещат изречения като „страхът стискаше гърлото, задушаваше гласа" [Тал66], а изрази като „буца в гърлото" или „пресъхнало гърло" са навлезли в разговорната реч като асоциации на „тъга". Тъй като изпитваната емоция влияе пряко на конфигурацията на вокалния тракт, бихме искали да извличаме характеристики, които описват тази конфигурация.

Да разгледаме по-подробно класическата постановка, показана на Фигура 2.1.1б, която описва цялостна система за производство на реч в по-опростен вариант. Речта, всъщност, представлява просто акустичната вълна, получена на края на системата - устни и ноздри - в следствие на изтласкания от белия дроб въздух.

Белият дроб работи като енергиен източник за тази система - въздушният поток, получен при свиването му от междуребрените мускули и диафрагмата, се пропагира нагоре по трахеята и през глотиса (отвора между гласните струни).
Действието на глотиса може да се види най-ясно при произнасяне на гласна. Гласните струни пропускат пропагирания въздух. Тъй като глотисът е стеснение, налягането в него в този момент е по-малко от това в който и да е от двата му края. Съгласно закона на Бернули, в някакъв момент то става толкова ниско, че позволява на гласните струни да се затворят. В следствие се натрупва налягане зад гласните струни заради тласкания от белия дроб въздух, което в някакъв момент ги принуждава да се отворят, и цикълът се повтаря отначало. В резултат се получава осцилиране на гласните струни. Честотата на отварянето и затварянето зависи от анатомични особености като еластичността и големината на гласните струни, налягането в белия дроб и други.
При мъжете тази честота е средно 125 Hz, а при жените - 210 Hz.
Акустичната вълна, която се получава в следствие на осцилацията, преминава през вокалния тракт, където се завихря при срещане на прегради като устни и зъби, и в крайна сметка напуска системата през някой от отворите.

При този процес се губи част от енергията поради различни фактори като съпротивлението на въздуха и поглъщането на вълната от меките и еластични стени на вокалния тракт.

В зависимост от начина, по който вълната напуска системата, можем да класифицираме произведените звуци по следния начин:

1. Озвучени

   При тези звуци гласните струни осцилират квази-периодично.

2. Проходни (фрикативни)

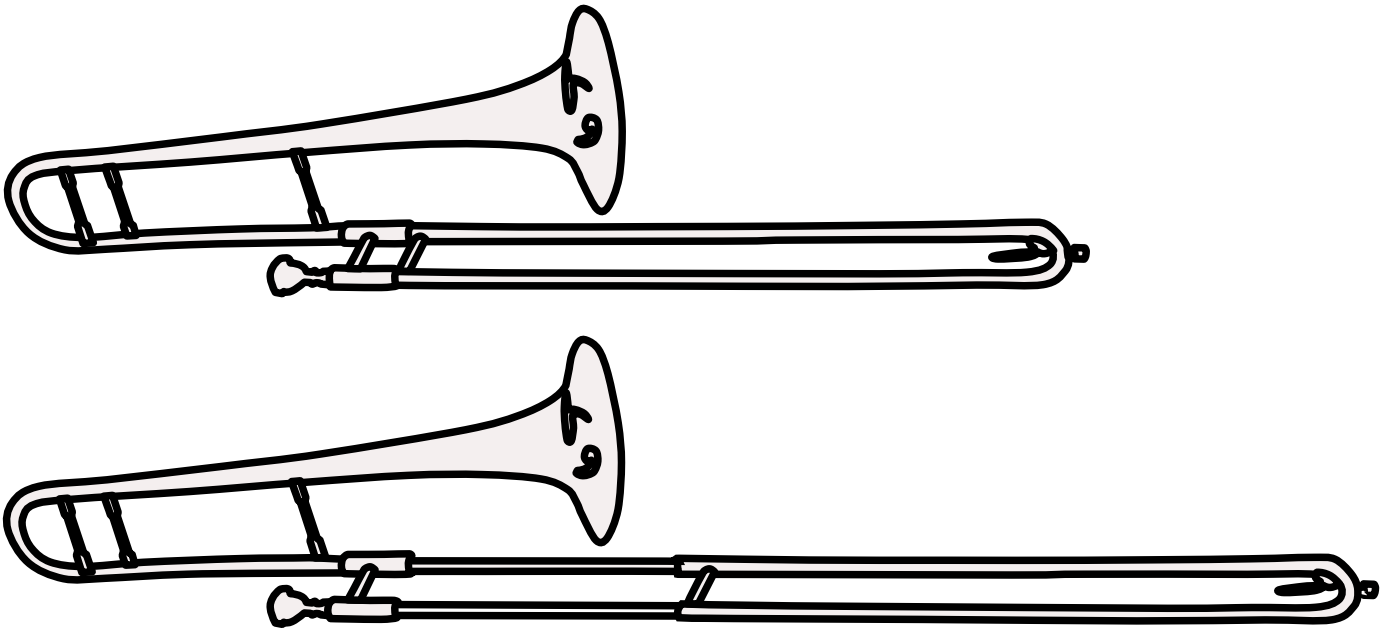
При образуването на проходни звуци, вълната среща преграда по пътя си (като например зъби, устни) и се получава турболенция, при опита да бъде избутан въздухът през преградата.

3. Преградни (експлозивни)

   Те се получават при напълно затворена преграда, зад която се натрупва налягане, което се освобождава рязко чрез отваряне на преградата.

Обикновено речта, която произнасяме, е разделена на думи, като отделните звукови единици в тях се наричат фонеми. За да се произнесе определена дума, вокалният тракт трябва да застане в правилната конфигурация за следващата фонема в думата. Когато вокалният тракт се наглася за дадена фонема, настъпват промени, като например стените на устната кухина се приближават или мекото небце, служещо като клапа към носната кухина, се затваря. Може да се усети, че при изговаряне на „а" отворът е много по-голям, отколкото при произнасяне на „у". Тази промяна влияе върху спектралните свойства на вокалния тракт.

Нека за улеснение си представим, че сме моделирали вокалния тракт с последователност от тръби, за да се абстрахираме от сложната му физическа структура. Тогава при смяна на фонемата, се променят дължината и диаметърът на тръбите. Това влияе на времето, за което акустичната вълна ще стигне до края на тръбата, и съответно на честотата, на която ще се получи резонанс. Тоест в зависимост от конфигурацията, ще се усилят или затихнат различни честоти, спрямо резонанса. Това свойство се нарича честотна пропускливост. Идеята лесно се вижда при свирене на духови инструменти.



Фигура 2.1.2: Тромбон

При тях по някакъв начин се променя изходът на вълната, например отпушване и запушване на дупки, и съответно честотата, на която се получава резонанс, тъй като пътят на вълната е скъсен или удължен. Както може да се види на Фигура 2.1.2, при тромбона буквално се сменя дължината на тръ-

бата, което означава, че на вълната ѝ трябва повече време, за да се отрази, тоест резонансът е на по-малка честота и съответно изходящият звук е по-нисък.

Това значи, че ако знаем как се пропагира вълната по отделните тръби на вокалния тракт и какви са спектралните свойства накрая, можем да съдим за текущата му конфигурация. В такъв случай, за да изследваме подлежащата емоция при реч е нужно да изследваме тези свойства в достатъчно кратък отрязък от време, в който конфигурацията е статична. Обикновено се приема, че този период е между 10 и 20 милисекунди ([RS78, стр. 98]).

В следващия раздел ще разгледаме как можем да моделираме вокалния тракт с модела на тръбите, за да можем да извлечем спектралните му свойства.

2.2 Приближаване с тръби

В тази глава ще разгледаме в неголяма дълбочина построяване на модел с тръби. За повече подробности, може да се проследи подробното изложение в [RS78] или по-сбитото в [Tay09].

За да извлечем спектралните свойства на вокалния тракт[1], трябва да моделираме системата за производство на реч. Освен това искаме да можем да отделим характеристиките на вокалния тракт от тези на останалите части на системата (като глотис и устни). Един такъв модел се получава с модела на тръбите, който ще бъде описан в този раздел.

За улеснение, нека разгледаме конкретна конфигурация. Например тази, при произнасянето на фонемата „ъ",тъй като е възможно най-проста. В този случай глотисът трепти, устата е отворена, а клапата към носната кухина е затворена.

Тъй като „ъ" е гласна, което е озвучен тип звук, глотисът g трепти псевдопериодично, после вълната преминава и се променя от вокалния тракт v и накрая излиза и се пречупва през устните r. Това означава, че ако глотисът има даден спектър, то вокалният тракт го променя (филтрира го), като усилва дадени честоти и заглушава други, до получаване на нов спектър. Устните допълнително филтрират спектъра. В крайна сметка получаваме нов сигнал, чийто спектър е резултат от умножението на спектрите на g, v и r.

Тоест, ако $g(t) \overset{\mathcal{FS}}{\longleftrightarrow} \mathcal{G}(z), v(t) \overset{\mathcal{FS}}{\longleftrightarrow} \mathcal{V}(z), r(t) \overset{\mathcal{FS}}{\longleftrightarrow} \mathcal{R}(z)$, а новият сигнал е y с $y(t) \overset{\mathcal{FS}}{\longleftrightarrow} Y(z)$, е изпълнено, че

$$\mathcal{Y}(z) = \mathcal{G}(z)\mathcal{V}(z)\mathcal{R}(z),$$

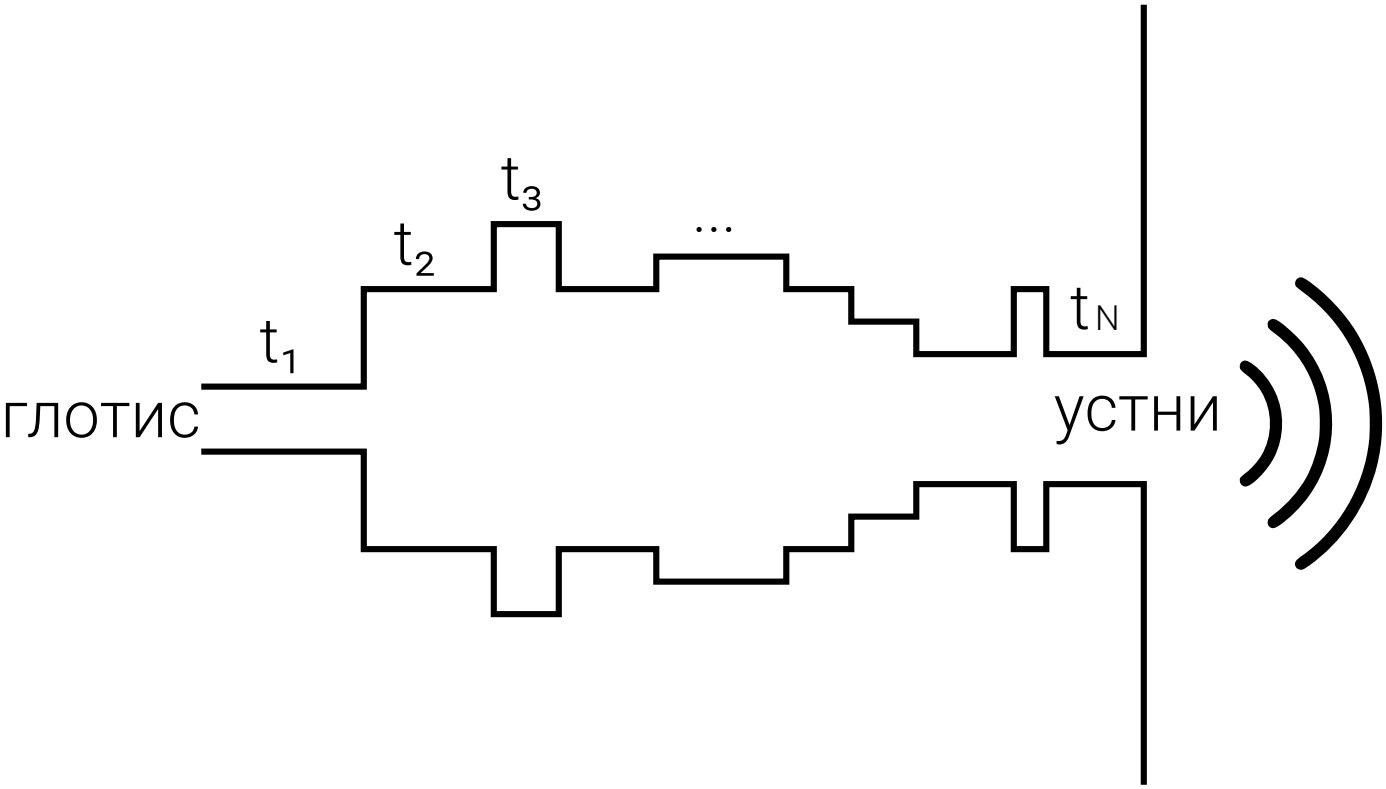
където $z = e^{i\omega_k}$ е прост сигнал с ъглова честота ω_k, а $\overset{\mathcal{FS}}{\longleftrightarrow}$ обозначава Фурие преобразуване, което е въведено в Приложение А.

[1] както се зарекохме в предния раздел

Във времевия домейн уравнението има вида $y(t) = g(t) * v(t) * r(t)$, както следва от Теорема за конволюцията за периодични дискретни сигнали също в Приложение А.

Тъй като в крайна сметка получаваме нов сигнал при устните, това, от което се интересуваме, са спектралните особености на $\mathcal{V}(z)\mathcal{R}(z)$, за които можем да си мислим като един общ филтър, описващ вокалния тракт. За да говорим за крайния сигнал y, трябва да знаем какво е действието на входния сигнал $\mathcal{G}$, който ще бъде променян от филтъра на вокалния тракт.

Бележка: Дефиницията за Фурие преобразуване изисква сигналът да е периодичен. В случая сме взели g да е такъв. Нека засега приемем, че сигналите v и r също са периодични. Това, а също и че периодът им съвпада с този на g, е показано в Свойство 1.



Фигура 2.2.1: Приближение на вокалния тракт с N тръби

По принцип стените на вокалния тракт са гладки и меки, но това се моделира трудно. Допълнително, формата му е сложна и специфична за всеки човек. Така че нека опростим ситуацията, като използваме приближение с N на брой тръби, номерирани $1...N$, с постоянно напречно сечение, както е показано на Фигура 2.2.1 За още по-голямо опростяване, нека няма и загуба на енергия, каквато би се получила по принцип.

Нека въведем следните стандартни означения:

1. c - скорост на звука в еластична среда
2. ρ - плътност на въздуха в тръбите
3. A - лицето на напречното сечение в тръба (константа)
4. $u = u(x,t)$ - е обемната скорост на позиция x в момента t
5. $p = p(x,t)$ - е звуковото налягане

Свойствата на звуковите вълни, преминаващи през течна среда в тръба, могат да се опишат с уравненията на Навие-Стокс:

$$-\frac{\partial \rho}{\partial x} = \frac{\rho}{A}\frac{\partial u}{\partial t} \tag{2.2.1a}$$

$$-\frac{\partial u}{\partial x} = \frac{A}{\rho c^2}\frac{\partial \rho}{\partial t} \tag{2.2.1б}$$

Тези уравнения са част от областта на динамиката на флуидите, която сериозно излиза извън обхвата на текущата дипломна работа. Затова, отново ще се позовем на [RS78], където е показано, че решенията на Уравнения 2.2.1 имат вида

$$u(x,t) = \left[u^+\left(t - \frac{x}{c}\right) - u^-\left(t + \frac{x}{c}\right)\right] \tag{2.2.2a}$$

$$p(x,t) = \frac{\rho c}{A}\left[u^+\left(t - \frac{x}{c}\right) + u^-\left(t + \frac{x}{c}\right)\right] \tag{2.2.2б}$$

Уравнение (2.2.2а) има следното значение: при преминаване от една тръба в друга, част от вълните ще преминат към следващата тръба, а част от тях ще се отразят наобратно. В такъв случай във всеки момент от време t и във всяка точка x на k-тата тръба, обемната скорост u ще зависи от обемната скорост на вълните, които вървят „напред", и тази на вълните, които вървят „назад". Вълните, които вървят „напред" и „назад", ще означаваме съответно с u^+ и u^-. Те са функции на времето и също измерват обемна скорост.

С помощта на тези две уравнения, ще изразим връзката между две съседни тръби.

2.2.1 Преминаване от една тръба в друга

За специфична тръба k, Уравнения (2.2.2) ще имат вида:

$$u_k(x,t) = \left[u_k^+\left(t - \frac{x}{c}\right) - u_k^-\left(t + \frac{x}{c}\right)\right] \tag{2.2.3a}$$

$$p_k(x,t) = \frac{\rho c}{A_k}\left[u_k^+\left(t - \frac{x}{c}\right) + u_k^-\left(t + \frac{x}{c}\right)\right], \tag{2.2.3б}$$

където A_k е лицето на напречното сечение на k-тата тръба, x е разстояние в нея ($0 \le x \le l_k$), t е момент от време.

Тъй като енергията трябва да се запази, въвеждаме допълнително условие за границата между две тръби:

$$u_k(l_k,t) = u_{k+1}(0,t) \tag{2.2.4a}$$

$$p_k(l_k,t) = p_{k+1}(0,t) \tag{2.2.4б}$$

Тук с l_k е дължината на k-тата тръба.

Когато заместим Уравнения (2.2.4) в (2.2.3), получаваме:

$$u_k^+\left(t-\frac{l_k}{c}\right)-u_k^-\left(t+\frac{l_k}{c}\right)=u_{k+1}^+(t)-u_{k+1}^-(t)$$

и

$$\frac{\rho c}{A_k}\left[u_k^+\left(t-\frac{l_k}{c}\right)+u_k^-\left(t+\frac{l_k}{c}\right)\right]=\frac{\rho c}{A_{k+1}}\left[u_{k+1}^+(t)+u_{k+1}^-(t)\right]$$

$$\Longleftrightarrow$$

$$\frac{A_{k+1}}{A_k}\left[u_k^+\left(t-\frac{l_k}{c}\right)+u_k^-\left(t+\frac{l_k}{c}\right)\right]=u_{k+1}^+(t)+u_{k+1}^-(t)$$

Нека означим с τ_k времето, за което вълна пропътува дължината на k-тата тръба, тоест $\tau_k=\frac{l_k}{c}$. Тогава имаме:

$$u_k^+(t-\tau_k)-u_k^-(t+\tau_k)=u_{k+1}^+(t)-u_{k+1}^-(t) \tag{2.2.5}$$

$$\frac{A_{k+1}}{A_k}[u_k^+(t-\tau_k)+u_k^-(t+\tau_k)]=u_{k+1}^+(t)+u_{k+1}^-(t) \tag{2.2.6}$$

Първо, нека да изразим обемната скорост на вълните, които вървят "напред" в $(k+1)$-та тръба (u_{k+1}^+), чрез тези, които са преминали от предната тръба (u_k^+) и тези, които се отразяват от текущата (u_{k+1}^-).

От (2.2.5) получаваме

$$u_k^-(t+\tau_k)=u_k^+(t-\tau_k)-u_{k+1}^+(t)+u_{k+1}^-(t) \tag{2.2.7}$$

Заместваме (2.2.7) в (2.2.6)

$$u_{k+1}^+(t)=\frac{A_{k+1}}{A_k}u_k^+(t-\tau_k)+\frac{A_{k+1}}{A_k}[u_k^+(t-\tau_k)-u_{k+1}^+(t)+u_{k+1}^-(t)]-u_{k+1}^-(t)$$

$$u_{k+1}^+(t)\left[1+\frac{A_{k+1}}{A_k}\right]=u_k^+(t-\tau_k)\frac{2A_{k+1}}{A_k}+u_{k+1}^-(t)\left[\frac{A_{k+1}}{A_k}-1\right]$$

$$u_{k+1}^+(t)\left[\frac{A_k+A_{k+1}}{A_k}\right]=u_k^+(t-\tau_k)\frac{2A_{k+1}}{A_k}+u_{k+1}^-(t)\left[\frac{A_{k+1}-A_k}{A_k}\right]$$

$$u_{k+1}^+(t)=u_k^+(t-\tau_k)\left[\frac{2A_{k+1}}{A_k+A_{k+1}}\right]+u_{k+1}^-(t)\left[\frac{A_{k+1}-A_k}{A_k+A_{k+1}}\right] \tag{2.2.8}$$

Коефициентът, който стои пред $u_k^+(t-\tau_k)$ в уравнение (2.2.8), представлява количеството енергия, която преминава от тръба k в тръба $k+1$, идваща от вълните, които се движат "напред" в k-тата тръба. Затова

$$t_k=\frac{2A_{k+1}}{A_k+A_{k+1}} \tag{2.2.9}$$

се нарича **коефициент на преминаване** за k-тия преход (преходът между тръби k и $k+1$).

Коефициентът пред $u^-_{k+1}(t)$ представлява количеството енергия, получено от вълните, които вървят "назад" в тръба $k+1$. Затова

$$r_k = \frac{A_{k+1} - A_k}{A_k + A_{k+1}} \tag{2.2.10}$$

се нарича **коефициент на отразяване** за k-тия преход.

Тъй като $A_k, A_{k+1} > 0, |A_{k+1} + A_k| > |A_{k+1} - A_k| \longleftrightarrow \frac{|A_{k+1} - A_k|}{|A_{k+1} + A_k|} < 1 \longleftrightarrow -1 < r_k < 1.$

Можем да забележим, че в специалния случай, в който напречните сечения на две съседни тръби са равни ($A_k = A_{k+1}$), би следвало всички вълни да преминават свободно. Наистина, ако заместим в уравнение (2.2.10), $r_k = 0$, а от (2.2.9) се вижда, че $t_k = 1$.

Нека изразим обемната скорост на вълните в тръба k чрез скоростта на вълните в $(k+1)$-вата тръба.

Първо разместваме уравнение (2.2.8)

$$u^+_k(t - \tau_k) = u^+_{k+1}(t)\left[\frac{A_k + A_{k+1}}{2A_{k+1}}\right] + u^-_{k+1}(t)\left[\frac{A_k - A_{k+1}}{2A_{k+1}}\right] \tag{2.2.11а}$$

Заместваме (2.2.11а) в (2.2.5)

$$u^-_k(t + \tau_k) = u^+_k(t - \tau_k) - u^+_{k+1}(t) + u^-_{k+1}(t)$$
$$u^-_k(t + \tau_k) = u^+_{k+1}(t)\left[\frac{A_k + A_{k+1}}{2A_{k+1}}\right] + u^-_{k+1}(t)\left[\frac{A_k - A_{k+1}}{2A_{k+1}}\right] - u^+_{k+1}(t) + u^-_{k+1}(t)$$
$$u^-_k(t + \tau_k) = u^+_{k+1}(t)\left[\frac{A_k + A_{k+1} - 2A_{k+1}}{2A_{k+1}}\right] + u^-_{k+1}(t)\left[\frac{A_k - A_{k+1} + 2A_{k+1}}{2A_{k+1}}\right]$$
$$u^-_k(t + \tau_k) = u^+_{k+1}(t)\left[\frac{A_k - A_{k+1}}{2A_{k+1}}\right] + u^-_{k+1}(t)\left[\frac{A_k + A_{k+1}}{2A_{k+1}}\right] \tag{2.2.11б}$$

Използвайки, че

$$\frac{1}{1 + r_k} = \frac{A_k + A_{k+1}}{A_{k+1} - A_k + A_{k+1} + A_k} = \frac{A_k + A_{k+1}}{2A_{k+1}}$$

$$\frac{r_k}{1 + r_k} = \frac{(A_{k+1} - A_k)}{(A_k + A_{k+1})}\frac{(A_k + A_{k+1})}{2A_{k+1}} = \frac{A_{k+1} - A_k}{2A_{k+1}},$$

и $-1 < r_k < 1, r_k \neq -1$ можем да запишем Уравнения (2.2.11) във вида:

$$u^+_k(t - \tau_k) = \frac{1}{1 + r_k}u^+_{k+1}(t) - \frac{r_k}{1 + r_k}u^-_{k+1}(t) \tag{2.2.12а}$$

$$u^-_k(t + \tau_k) = -\frac{r_k}{1 + r_k}u^+_{k+1}(t) + \frac{1}{1 + r_k}u^-_{k+1}(t) \tag{2.2.12б}$$

Сега да разгледаме Уравнения (2.2.12) в честотния домейн. Избираме $z = e^{i\omega_k}$ и използваме свойствата, описани в Приложение А. Тоест, че $u_k[t-\tau_k] \overset{\mathcal{FS}}{\longleftrightarrow} z^{-\tau_k} U_k(z)$

$$z^{-\tau_k} U_k^+(z) = \frac{1}{1+r_k} U_{k+1}^+(z) - \frac{r_k}{1+r_k} U_{k+1}^-(z)$$

$$z^{\tau_k} U_k^-(z) = -\frac{r_k}{1+r_k} U_{k+1}^+(z) + \frac{1}{1+r_k} U_{k+1}^-(z)$$

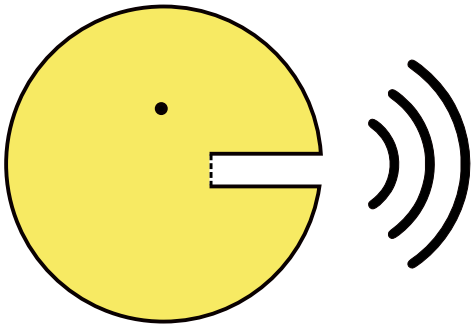
$$\iff \tag{2.2.13а}$$

$$U_k^+(z) = \frac{z^{\tau_k}}{1+r_k} U_{k+1}^+(z) - \frac{r_k z^{\tau_k}}{1+r_k} U_{k+1}^-(z) \tag{2.2.13б}$$

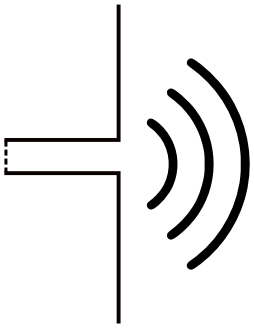
$$U_k^-(z) = -\frac{r_k z^{-\tau_k}}{1+r_k} U_{k+1}^+(z) + \frac{z^{-\tau_k}}{1+r_k} U_{k+1}^-(z) \tag{2.2.13в}$$

По този начин получихме връзката между две съседни тръби. За да получим общия модел, трябва да отчетем двете специални ситуации - при първата и при последната тръба.

2.2.2 Ограничения при устните



(а) Представяне на устните като отвор в сферична преграда



(б) Представяне на устните като отвор в безкрайна равнина

Фигура 2.2.2: Представяне на устните като отвор в преграда

Един разумен начин да представим изхода при устните е показан на Фигура 2.2.2а. На фигурата се вижда как звуковите вълни, които напускат системата, претърпяват дифракция при отвора в сферичната повърхност, моделираща главата. Представянето на тази дифракция е сложно, затова ще се опитаме да го опростим.

Ако отворът на устните е много малък спрямо размера на сферата, то можем да си мислим за преградата като за безкрайна равнина, както е показано на Фигура 2.2.2б

Използвайки тези предположения, в [RS78] е показано, че съществува следната връзка между налягането и обемната скорост:

$$\mathcal{P}_N(l_N, z) = Z_L(z) \mathcal{U}_N(l_N, z), \tag{2.2.14}$$

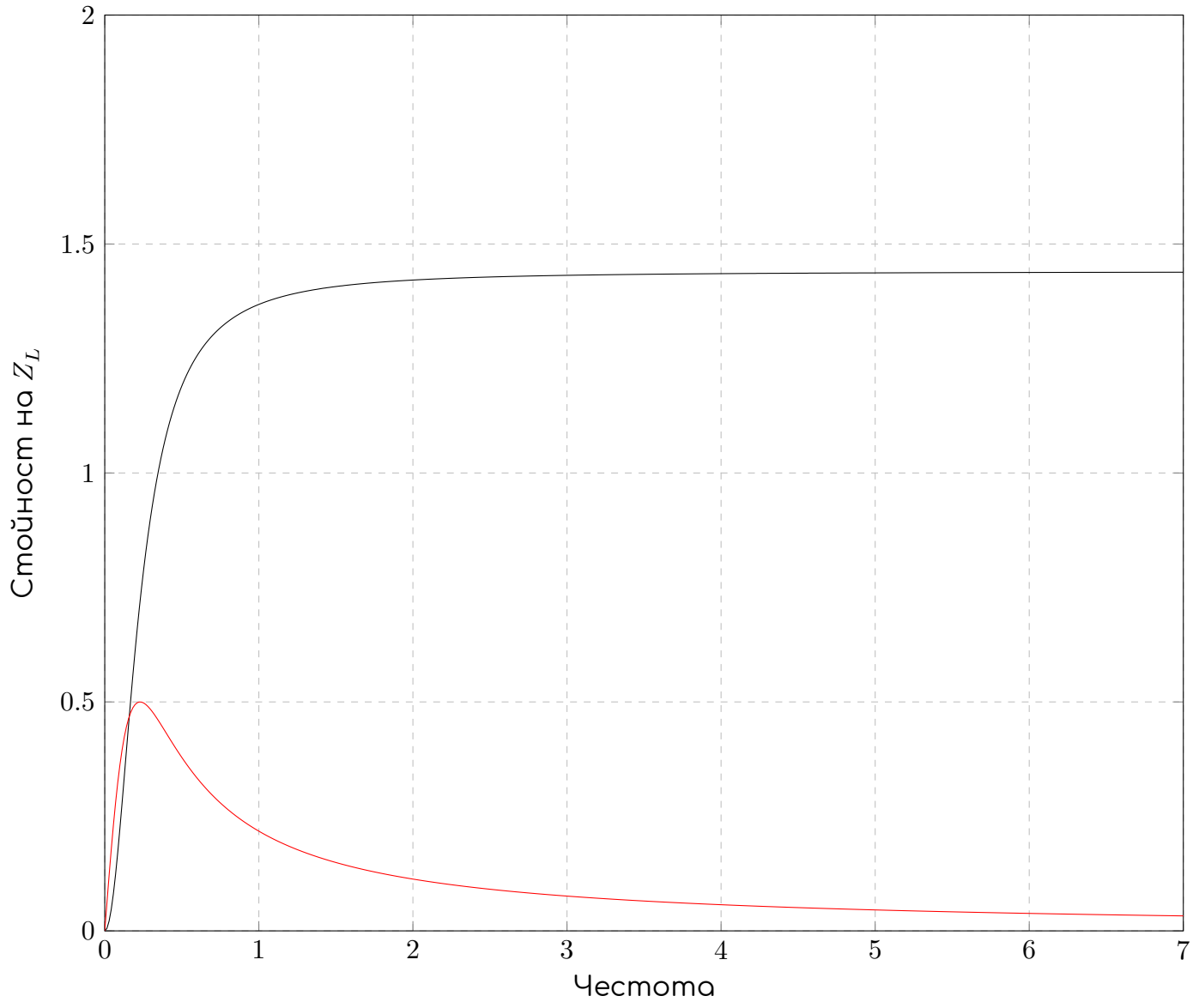
където $Z_L(z)$ се нарича радиационен импеданс (пълно съпротивление), описва загубите, които се получават на изхода, и има вида:

$$Z_L(z) = \frac{i\omega L_r R_r}{R_r + i\omega L_r}, \tag{2.2.15}$$

Където $z = e^{i\omega}$ означава сигнал с ъглова честота ω, а L_r и R_r са константи, определени от размера на отвора на устата. За практически цели се избират:

$$R_r = \frac{128}{9\pi^2} \approx 1.44 \qquad\qquad L_r = \frac{8a}{3\pi c}$$

a - радиус на отвора, c - скоростта на звука.



Фигура 2.2.3: Реална(черно) и имагинерна(червено) част на Z_L

От Фигура 2.2.3 се вижда, че при много ниски честоти $Z_L(z) \approx 0$, което значи, че съпротивлението на устните е почти нулево. При средни честоти ($\omega L_r \ll R_r$),

$Z_L(z) \approx i\omega L_r$, а високи честоти, ($\omega L_r \gg R_r$) $Z_L(z) \approx R_r$. Това значи, че загубите при устните са най-големи при високи честоти, тъй като тогава импедансът е най-голям.

Ако предположим, че честотата z е висока, $Z_L \approx R_r$ е реално число и не зависи от z, тоест $Z_L(z) = Z_L$.

Тогава ако $p_N(l_N, t) \overset{\mathcal{FS}}{\longleftrightarrow} P_N(l_N, z), u_N(l_N, t) \overset{\mathcal{FS}}{\longleftrightarrow} U_N(l_N, z)$ и съответно предположим, чеZ_L е Z_L, можем да разгледаме уравнението (2.2.14) във времевия домейн:

$$p_N(l_N, t) = Z_L u_N(l_N, t),$$

Ако използваме Уравнения (2.2.3) и заместим с $\tau_N = \frac{l_N}{c}$, получаваме:

$$\begin{aligned}
&\frac{\rho c}{A_N}\left[u_N^+(t-\tau_N) + u_N^-(t+\tau_N)\right] = Z_L\left[u_N^+(t-\tau_N) - u_N^-(t+\tau_N)\right] \\
&u_N^-(t+\tau_N)\frac{(\rho c + A_N Z_L)}{A_N} = u_N^+(t-\tau_N)\frac{(A_N Z_L - \rho c)}{A_N} \\
&u_N^-(t+\tau_N) = -r_L u_N^+(t-\tau_N), \text{ където}
\end{aligned} \tag{2.2.16}$$

$$r_L = \left(\frac{\frac{\rho c}{Z_L} - A_N}{\frac{\rho c}{Z_L} + A_N}\right)$$

В случая, в който $Z_L \approx i\omega L_r$ е комплексно, в [RS78] се показва, че уравнение (2.2.16) остава в сила, но в този случай и r_L също ще бъде комплексно.

2.2.3 Ограничения при глотиса

Както при устните, така и при глотиса, трябва да се отчете импедансът. Например когато глотисът е затворен, импедансът е безкраен, а обемната скорост нулева.

Връзката $U_1(0, z) = U_G(z)$ е твърде наивна и в [RS78] е показано, че по-добро приближение би било:

$$U_1(0, z) = U_G(z) - \frac{P_1(0, z)}{Z_G(z)}, \tag{2.2.17}$$

където $z = e^{i\omega}$, Z_G описва импеданса на глотиса и $Z_G(z) = R_G + i\omega L_G$,

L_G, R_G - константи.

Отново предполагайки, че Z_G е реално, тоест $\omega_k L_G << R_G$, можем да разгледаме уравнение (2.2.17) във времевия домейн.

Нека $u_1(0, t) \overset{\mathcal{FS}}{\longleftrightarrow} U_1(0, z), p_1(0, t) \overset{\mathcal{FS}}{\longleftrightarrow} P_q(0, z)$ за фиксиран първи аргумент и съответно Z_G е константа и $Z_G \overset{\mathcal{FS}}{\longleftrightarrow} Z_G(z) = Z_G$:

$$u_1(0,t) = u_G(t) - \frac{p_1(0,t)}{Z_G}$$

Ако използваме Уравнения (2.2.3), получаваме

$$u_1^+(t) - u_1^-(t) = u_G(t) - \frac{\rho c}{A_1}\left[\frac{u_1^+(t) + u_1^-(t)}{Z_G}\right]$$
$$u_1^+(t)\left[1 + \frac{\rho c}{A_1 Z_G}\right] = u_G(t) + u_1^-(t)\left[1 - \frac{\rho c}{A_1 Z_G}\right]$$
$$u_1^+(t) = u_G(t)\left[\frac{A_1 Z_G}{A_1 Z_G + \rho c}\right] + u_1^-(t)\left[\frac{A_1 Z_G - \rho c}{A_1 Z_G + \rho c}\right]$$

$$u_1^+(t) = u_G(t)\left[\frac{1 + r_G}{2}\right] + r_G u_1^-(t) \tag{2.2.18}$$

където $r_G = \left(\frac{A_1 Z_G - \rho c}{A_1 Z_G + \rho c}\right)$ и е изпълнено

$$\frac{1 + r_G}{2} = \frac{A_1 Z_G + \rho c + A_1 Z_G - \rho c}{2(A_1 Z_G + \rho c)} = \frac{A_1 Z_G}{A_1 Z_G + \rho c}$$

Ако се върнем в честотния домейн:

$$U_G(z) = \left[\frac{2}{1 + r_G}\right] U_1^+(z) - \left[\frac{2 r_G}{1 + r_G}\right] U_1^-(z), \tag{2.2.19}$$

Отново в [RS78] е показано, че ако Z_G е комплексно, уравнението (2.2.18) е в сила и в този случай r_G също е комплексно. За улеснение обикновено Z_L и Z_G се взимат реални.

2.2.4 Общ вид на $\mathcal{V}$

За да видим общия вид на $\mathcal{V}$, нека засега всички тръби имат равна дължина и тя е $\tau_i = \frac{1}{2}, i \in [1...N]$ Тогава уравнения (2.2.13) имат вида:

$$U_k^+(z) = \frac{z^{1/2}}{1 + r_k} U_{k+1}^+(z) - \frac{r_k z^{1/2}}{1 + r_k} U_{k+1}^-(z) \tag{2.2.20а}$$

$$U_k^-(z) = -\frac{r_k z^{-1/2}}{1 + r_k} U_{k+1}^+(z) + \frac{z^{-1/2}}{1 + r_k} U_{k+1}^-(z) \tag{2.2.20б}$$

За да опишем граничните условия при устните, дефинираме $U_{N+1}(z)$ да е Фурие трансформацията на входа на несъществуваща $(N+1)$ тръба. Тази тръба е безкрайно дълга и заради това скоростта на вървящите „назад" вълни трябва да бъде 0

Тоест дефинираме:

$$U^+_{N+1}(z) = U_L(z) \qquad (2.2.21)$$
$$U^-_{N+1}(z) = 0$$

Също така искаме коефициентът на отражение на последната истинска тръба да е равен на коефициент на отражение при устните, а именно $r_N = r_L$

$$\left(\frac{A_{N+1} - A_N}{A_{N+1} + A_N}\right) = \left(\frac{\frac{\rho c}{Z_L} - A_N}{\frac{\rho c}{Z_L} + A_N}\right)$$

Това ни дава, че $A_{N+1} = \frac{\rho c}{Z_L}$

Ако представим Уравнения (2.2.20) в матричен вид, получаваме:

$U_k = Q_k U_{k+1}$ за

$$U_k = \begin{bmatrix} U^+_k(z) \\ U^-_k(z) \end{bmatrix} \qquad Q_k = \begin{bmatrix} \frac{z^{1/2}}{1+r_k} & \frac{-r_k z^{1/2}}{1+r_k} \\ \frac{-r_k z^{-1/2}}{1+r_k} & \frac{z^{-1/2}}{1+r_k} \end{bmatrix}$$

$$U_1 = Q_1 U_2 = Q_1 Q_2 U_3 = \cdots = Q_1 ... Q_N U_{N+1} = \left[\prod_{i=1}^{N} Q_i\right] U_{N+1}$$

За специалното ограничение при глотиса, разглеждаме матричния вид на Уравнение 2.2.19

$$U_G(z) = \begin{bmatrix} \frac{2}{1+r_G}, & -\frac{2r_G}{1+r_G} \end{bmatrix} U_1$$

Ограниченията (2.2.21) за U_L ни дават, че

$$U_{N+1} = \begin{bmatrix} U^+_{N+1}(z) \\ U^-_{N+1}(z) \end{bmatrix} = \begin{bmatrix} U_L(z) \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix} U_L(z)$$

В [RS78] е описано, че обемната скорост при устните може да се получи от обемната скорост при глотиса, умножена по филтъра на вокалния тракт в честотния домейн, тоест:

$U_L(z) = U_G(z)V(z)$.

Тогава, ако заместим, получаваме:

$$\frac{1}{\mathcal{V}(z)} = \frac{U_G(z)}{U_L(z)} = \frac{\left[\begin{array}{cc} \frac{2}{1+r_G}, & -\frac{2r_G}{1+r_G} \end{array}\right] \prod_{i=1}^{N} Q_i \left[\begin{array}{c} 1 \\ 0 \end{array}\right] U_L(z)}{U_L(z)} =$$

$$= \left[\begin{array}{cc} \frac{2}{1+r_G}, & -\frac{2r_G}{1+r_G} \end{array}\right] \prod_{i=1}^{N} Q_i \left[\begin{array}{c} 1 \\ 0 \end{array}\right] \tag{2.2.22}$$

Нека изразим Q_k по следния начин:

$$Q_k = \left[\begin{array}{cc} \frac{z^{1/2}}{1+r_k} & \frac{-r_k z^{1/2}}{1+r_k} \\ \frac{-r_k z^{-1/2}}{1+r_k} & \frac{z^{-1/2}}{1+r_k} \end{array}\right] = z^{1/2} \left[\begin{array}{cc} \frac{1}{1+r_k} & \frac{-r_k}{1+r_k} \\ \frac{-r_k z^{-1}}{1+r_k} & \frac{z^{-1}}{1+r_k} \end{array}\right] = z^{1/2}\hat{Q}_k$$

Тогава уравнение (2.2.22) има вида

$$\frac{1}{\mathcal{V}(z)} = \frac{U_G(z)}{U_L(z)} = z^{N/2} \left[\begin{array}{cc} \frac{2}{1+r_G}, & -\frac{2r_G}{1+r_G} \end{array}\right] \prod_{i=1}^{N} \hat{Q}_i \left[\begin{array}{c} 1 \\ 0 \end{array}\right] \tag{2.2.23}$$

Нека изразим общия вид на $\mathcal{V}(z)$. За $N = 2$, например, има вида:

$$\mathcal{V}(z) = \frac{0.5(1+r_G)\prod_{i=1}^{2}(1+r_i)z^{-1}}{1+(r_G r_1 + r_1 r_2)z^{-1} + (r_G r_2)z^{-2}}, \tag{2.2.24}$$

както е показано в Пример 3 на Приложение към Сигнал от реч.

Може да се покаже, че в общия случай за произволно N, Уравнение 2.2.23 може да се развие итеративно до:

$$\mathcal{V}(z) = \frac{0.5(1+r_G)\prod_{i=1}^{N}(1+r_i)z^{-N/2}}{1-\sum_{i=1}^{N}\alpha_i z^{-i}} \tag{2.2.25}$$

Както се вижда от Пример 2, предположението, че $\tau_i = \frac{1}{2}$ е разумно, тъй като при тази стойност се получава най-голяма изразителна сила, както е отбелязано и в Бележката.

2.2.5 Общ вид на $\mathcal{R}$

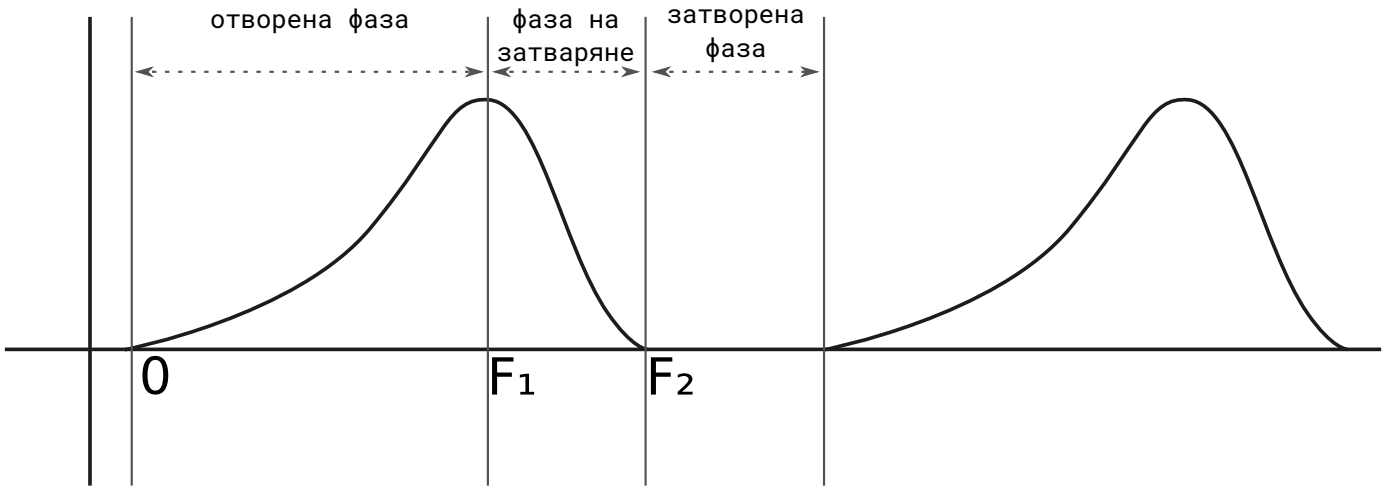
Моделът, описан в Подраздел 2.2.2, се моделира прекалено трудно. Може да се покаже ([Tay09],[Qua01]) , че ефектът от радиацията се приближава достатъчно добре с една нула в единичната окръжност, тоест:

$$\mathcal{R}(z) = (1 - \gamma z^{-1}), \gamma < 1 \tag{2.2.26}$$

Обикновено $\gamma \approx 0.97$.

2.2.6 Общ вид на $\mathcal{G}$

За да се симулира действието на глотиса, трябва да отчетем как се държи той при изговаряне на различни видове звуци.



Фигура 2.2.4: Пример за импулс от глотиса

В случая на гласна, както бяхме приели за улеснение, той трепти периодично и видът му може да се види на Фигура 2.2.4. Можем да разделим този сигнал на три фази:

1. Отворена фаза с край F_1
2. Фаза на затваряне с начало F_1 и край F_2
3. Затворена фаза с начало F_2

Това може да се опише като функция на времето по следния начин.

$$g(t) = \begin{cases} \frac{1}{2}(1 - cos\,(\pi n/F1)), & 0 \leq t \leq F_1 \\ cos\,(\pi(n - F_1)/2(F_2 - F_1)), & F_1 \leq t \leq F_2 \\ 0, & \text{иначе} \end{cases}$$

Поведение, като това на $\mathcal{G}$, може да се приближава с два полюса, както е показано в Приложение за полюси и нули, тоест

$$\tilde{\mathcal{G}}(z) = \frac{1}{(1 - \alpha_1 z^{-1})(1 - \beta z^{-1})},$$

но по този начин не се отчита самата форма на сигнала.

В [Qua01] е показано, че по-добро приближение се получава при $\alpha = \beta$ и $\mathcal{G}(z) = \tilde{\mathcal{G}}(-z)$, тоест

$$\mathcal{G}(z) = \frac{1}{(1-\beta z)^2}, \beta < 1$$

Сигналът, показан на Фигура 2.2.4, е почти идеален. В действителност е почти невъзможно да се поддържа тон, който има еднакви разстояния между пиковете и еднаква амплитуда. Отклонението от истинския период се нарича джитер[2]. Другият ефект, който е важен за истинския човешки глас, е трептенето[3], тоест разликата в амплитудите. Освен за естествеността на гласа, тези характеристики могат да носят информация и за емоционалното състояние. Висок джитер може да означава дрезгав глас, но също може да се предизвика при чувство на стрес или страх. Включването им в модела се постига с допълнителни нули, което ни дава и крайния вид на $\mathcal{G}$ в случая на озвучен звук:

$$\mathcal{G}(z) = \frac{\prod_{i=0}^{K}(1-\beta_i z^{-1})}{(1-\beta z)^2} \tag{2.2.27}$$

Да разгледаме случая с беззвучен звук. При съгласните например сигналът от глотиса е случайна редица с плосък спектър (тоест има почти еднаква мощност в целия спектър). Добър начин да се моделира е чрез генератор на бял шум.

2.2.7 Общ вид на $\mathcal{Y}$

От уравнения 2.2.25, 2.2.26, 2.2.27 следва, че видът на $\mathcal{H}$ е

$$\mathcal{Y}(z) = \mathcal{G}(z)\mathcal{V}(z)\mathcal{R}(z) = \left[\frac{\prod_{i=0}^{K}(1-\beta_i z^{-1})}{(1-\beta z)^2}\right]\left[\frac{0.5(1+r_G)\prod_{i=1}^{N}(1+r_i)z^{-N/2}}{1-\sum_{i=1}^{N}\alpha_i z^{-i}}\right][(1-\gamma z^{-1})]$$

При определени стойности на коефициентите, видът на $\mathcal{Y}$ може да се запише като:

$$\mathcal{Y}(z) = \mathcal{G}(z)\frac{\sum_{m=0}^{M} b_m z^{-m}}{\sum_{k=0}^{K} a_k z^{-k}}, \tag{2.2.28}$$

[2]На английски jitter, думата е заемка, тъй като няма друг възприет български термин. Руският термин е джиттер

[3]На английски shimmer. Тук също не мога да намеря подходящ български термин, но в случая и единственият руски, който открих, е шиммер

от което ще се възползваме в следващия раздел.

2.3 Представяне със системи

Нека имаме чистия сигнал от глотиса $g[t]$. При преминаването му през вокалния тракт и устните, той се променя, в следствие на различни фактори като турболенция, поглъщане, отразяване, в следствие на което на изхода (устните), получаваме сигнала $y[n]$.
Дефиниция. (Система)
Механизъм, който манипулира един или повече сигнали с някаква цел до получаване на нов сигнал, се нарича система.

Обикновено в практическия свят се използват системи, чието действие е предварително известно (и желано). Такива системи наричаме **филтри**. Филтрите обикновено изпълняват някаква точно определена манипулация върху сигнала, например да премахват всички честоти под или над определена честота.

С $g[n] \mapsto y[n]$ ще бележим, че y е отговорът на системата за вход g. В такъв случай системата, която ще разгледаме, е тази на вокалния тракт. Ще ни интересуват следните няколко класа системи.
Дефиниция. (Линейна система)
Ако $x_1[n] \mapsto y_1[n]$ и $x_2[n] \mapsto y_2[n]$, то системата е линейна $\longleftrightarrow$

$\forall a, b \in \mathbb{R}\,(ax_1[n] + bx_2[n] \mapsto ay_1[n] + by_2[n])$
Дефиниция. (Времево-инвариантна система)
Нека $x[n] \mapsto y[n]$. Тогава, ако за всяко $n_0 : x[n - n_0] \mapsto y[n - n_0]$, то системата е времево-инвариантна.
Свойство 1. Ако системата е времево-инвариантна и сигналът x е периодичен с период N, то и изходът на системата y е периодичен с период N:

$x[n] \mapsto y[n]$ и $x[n] = x[n + N] \implies x[n + N] \mapsto y[n]$. Но от времевата инвариантност знаем, че $x[n + N] \mapsto y[n + N] \implies y[n] = y[n + N]$

Специален подклас на линейните, времево-инвариантни системи, е класът на системите, удовлетворяващи диференчното уравнение от ред N с константни коефициенти:

$$\sum_{k=0}^{N} a_k y[n-k] = \sum_{m=0}^{M} b_m x[n-m] \tag{2.3.1}$$

Вокалният тракт е времево-инвариантна система, защото изходът $y[n]$ не зависи от момента от време, а само от специфичната му конфигурация в текущия момент, т.е. положението на езика, устните, зъбите. Нека предположим, че вокалният тракт е линейна, времево-инвариантна система, която удовлетворява уравнение Уравнение 2.3.1, и да разгледаме свойствата.

Искаме да опишем как работи тази система. За момента знаем как ще реагира тя, ако ѝ подадем входен сигнал $g[n]$. Но вместо да разглеждаме отговора

на системата за широк спектър от входни функции, ще е полезно да имаме характеризация, която не зависи от входа.

Първо да разгледаме входа по различен начин. Ако за всеки момент от време n_0 имаме импулси със сила $g[n_0]$, то можем да мислим за входния сигнал $g[n]$ като за сума от тези импулси. Тоест, нека имаме дискретния единичен импулс:

$$\delta[n] = \begin{cases} 1, & n = 0 \\ 0, & \text{иначе} \end{cases}$$

Тогава можем да представим входния сигнал $g[n]$ като

$$g[n] = \sum_{k=-\infty}^{\infty} g[k]\delta[n-k]$$

Нека $\delta[n-k] \mapsto h_k[n]$. Тъй като системата е линейна, е изпълнено, че:

$$g[n] = \sum_{k=-\infty}^{\infty} g[k]\delta[n-k] \mapsto \sum_{k=-\infty}^{\infty} g[k]h_k[n] = y[n] \tag{2.3.2}$$

Времевата инвариантност ни казва, че ако $\delta[n] \mapsto h[n]$, то $\delta[n-k] \mapsto h[n-k]$, следователно в случая на вокалния тракт Уравнение 2.3.2 има вида:

$$y[n] = \sum_{k=-\infty}^{\infty} g[k]h_k[n] = \sum_{k=-\infty}^{\infty} g[k]h[n-k] \quad , \tag{2.3.3}$$

или записано като конволюция $y[n] = (g * h)[n]$.

Ако разгледаме Фурие преобразуванията на y, g, h, които са съответно $\mathcal{Y}, \mathcal{G}, \mathcal{H}$, в $z = e^{iw_k}$, получаваме:

$$\mathcal{Y}(z) = \mathcal{G}(z)\mathcal{H}(z)$$
$$\mathcal{H}(z) = \frac{\mathcal{Y}(z)}{\mathcal{G}(z)} \tag{2.3.4}$$

$\mathcal{H}$ се нарича предавателна функция за системата.

Да разгледаме фурие преобразуването на Уравнение 2.3.1 за входен сигнал g.

$$\left[\sum_{k=0}^{N} a_k z^{-k}\right] \mathcal{Y}(z) = \left[\sum_{m=0}^{M} b_m z^{-m}\right] \mathcal{G}(z)$$
$$\frac{\mathcal{Y}(z)}{\mathcal{G}(z)} = \frac{\sum_{m=0}^{M} b_m z^{-m}}{\sum_{k=0}^{N} a_k z^{-k}} \tag{2.3.5}$$

Когато заместим Уравнение 2.3.5 в Уравнение 2.3.4, получаваме

$$\mathcal{H}(z) = \frac{\sum_{m=0}^{M} b_m z^{-m}}{\sum_{k=0}^{N} a_k z^{-k}} \tag{2.3.6}$$

В Раздел 2.2 видяхме, че резултатния сигнал y, който се получава при изходите на системата, има следния вид:

$$\mathcal{Y}(z) = \mathcal{G}(z)\mathcal{V}(z)\mathcal{R}(z) = \mathcal{G}(z)\frac{\sum_{m=0}^{M} b_m z^{-m}}{\sum_{k=0}^{K} a_k z^{-k}} \tag{2.2.28}$$

Това означава, че $\mathcal{V}(z)\mathcal{R}(z)$, всъщност описват предавателната функция на системата $g[n] \mapsto y[n]$, тоест $\mathcal{H}(z) = \mathcal{V}(z)\mathcal{R}(z)$.

Следователно, производството на реч се описва от системата $\mathcal{Y}(z) = \mathcal{G}(z)\mathcal{H}(z)$, а $\mathcal{H}$ съдържа информацията за вокалния тракт. Характеристиките, които ще изберем, трябва да носят тази информация за вокалния тракт, тоест трябва да отделят входния сигнал $\mathcal{G}$ от филтъра $\mathcal{H}$, извличайки информацията за подлежащата емоция, която се надяваме, че е кодирана в $\mathcal{H}$.

Изборът на характеристики е описан по-подробно в следващия раздел.

2.4 Характеристики

2.4.1 Избор

На дневен ред е избирането на характеристики, които отразяват идеята за разделяне на информацията за вокалния тракт $\mathcal{H}$ от входния сигнал $\mathcal{G}$. Имаме $\mathcal{Y}(z) = \mathcal{G}(z)\mathcal{H}(z)$.

Нека вземем логаритъм от модула:

$$log(|\mathcal{Y}(z)|) = log(|\mathcal{G}(z)|) + log(|\mathcal{H}(z)|)$$

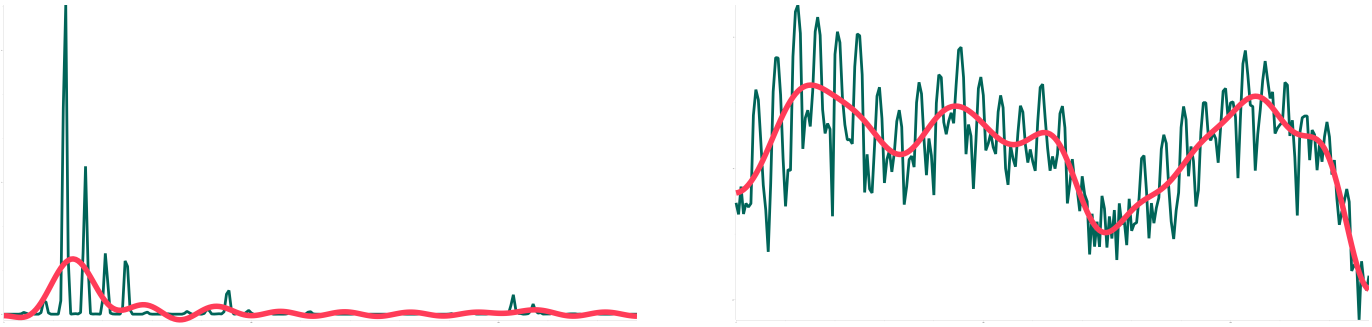
Обратното Фурие преобразуване ни дава вид във времевия домейн:

$$c_y[n] = c_g[n] + c_h[n]$$

Сега вече имаме сбор на входния сигнал и този на филтъра, вместо конволюция във времевия домейн.

Да видим каква е идеята зад тези преобразувания. Имаме, че $\mathcal{G}(z)$ и $\mathcal{H}(z)$ са комплексни числа и взимайки модула им, губим информация за фазата. Това не

е проблем, тъй като човешкото ухо не е особено чувствително към нея, затова обикновено ни трябва само амплитудата (и следователно модула).



(а) Линейна скала без логаритъм от модула (б) Линейна скала с логаритъм от модула

Фигура 2.4.1: Графика на спектъра на сигнал, получен при произнасяне на "а-а-а-а-а"

Взимането на логаритъм от модула цели да подчертае периодичността на сигнала, идващ от глотиса.

На Фигура 2.4.1а се виждат пиковете, породени от фундаменталната честота, на която трепти глотиса, и хармоничните ѝ честоти. Поради загубата на енергията в системата за производство на реч, не всички хармонични честоти имат една и съща амплитуда. На Фигура 2.4.1б се вижда, че взимането на логаритъм от модула помага за изравняването на хармоничните амплитуди и кара графиката да изглежда "по-периодична". Сега нека разгледаме логаритмувания спектър като сигнал и му направим Фурие преобразуване до получаване на така наречения кепстър. Тъй като разглеждаме спектъра като сигнал, наличието на периодични амплитуди ще се преведе до пик в получения кепструм на честотата, отговаряща (горе-долу, имайки предвид джитера) на основната честота на глотиса. Информацията, която описва вокалния тракт, е с много по-малко изразена периодичност в спектъра, затова ще се запази в ниските честоти на кепстъра. Тоест $c_g[n]$ ще са коефициентите в кепстъра на високите честоти, а $c_h[n]$ в ниските. За практически цели обикновено се избират първите 13 коефициента на кепстъра. Тези коефициенти се наричат Mel Frequency Cepstral Coefficients (MFCC), където Mel скалата е логаритмична скала. Повече детайли за извличането им са описани в следващия подраздел.

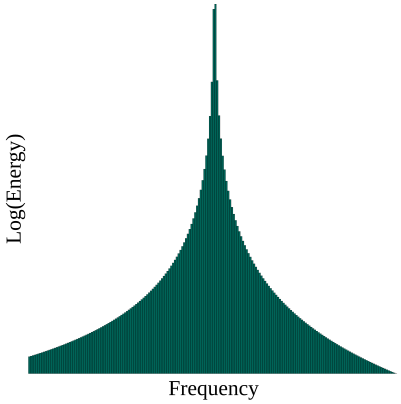
2.4.2 Извличане

Ще извличаме характеристики от подаден аудио файл в wav формат. Първо, съдържанието на файла се прочита в масив от 64-битови float числа. Елементите на този масив се наричат дискрети (samples). Броят им зависи от честотата на дискретизация (тоест колко измервания са направени за една секунда), която обикновено е 16kHz, 44100Hz или 48kHz. Броят на дискретите определя броя на коефициентите на Фурие преобразуването. Тъй като

сигналът е реален, от свойствата следва, че максималната честота, която можем да измерим, е честотата на Найкуист, равна на броя дискрети за секунда върху две. Тоест - половината на честотата на дискретизация. Колкото е по-голяма Найкуист честотата, толкова по-добра честотна резолюция получаваме.
Базирайки се на идеята, че вокалният тракт е статичен за кратък период от време, накъсваме масива на отделни застъпващи се парчета - фреймове - в рамките на които сигналът е статичен. За да се определи дължината на фрейма, трябва да се вземат предвид две неща: от една страна колкото повече дискрети имаме, толкова по-добра честотна резолюция получаваме. От друга, колкото повече дискрети взимаме, толкова по-голям е шансът да се смени конфигурацията на вокалния тракт. За да се справим с този дуализъм, компромисните стойности, които са избрани в описваната имплементация, са 25 милисекунди за дължина на фрейм и 10 милисекунди за разстояние между два последователни фрейма. Един фрейм описва една конфигурация на вокалния тракт. Целим да извлечем MFCC коефициенти за всеки фрейм. Това означава, че трябва да се направи Фурие преобразуване, което изисква сигналът да е периодичен, а данните във фреймовете не са. За тази цел, всеки фрейм се умножава по специално избрана функция, наречена прозорец[4]. Тази функция е нула навсякъде, освен в избран интервал, и обикновено е симетрична около средата на този интервал. Когато фреймът се умножи по прозорец със същата дължина, стойностите в краищата се нулират. Това прави полученият сигнал периодичен с период дължината на фрейма, тъй като започва в нула и завършва отново в нула.

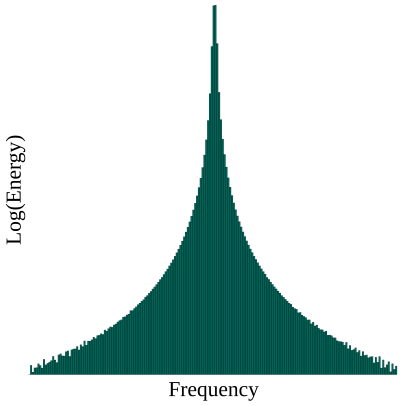
[4]Прозорецът е моята врата и аз вървя към тях и ги разпитвам.

(а) Действие на правоъгълен прозорец върху синусоида



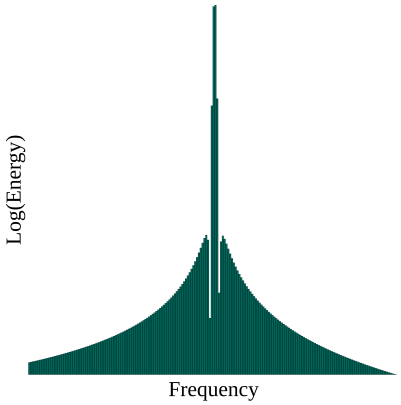
(б) Спектър на синусоида с правоъгълен прозорец

(в) Действие на прозорец на Хан върху синусоида



(г) Спектър на синусоида с прозорец на Хан

(д) Действие на прозорец на Хеминг върху синусоида



(е) Спектър на синусоида с прозорец на Хеминг

Фигура 2.4.2: Действие на често срещани прозоречни функции

Най-простият прозорец, който постига желания ефект, е правоъгълният прозорец с интервал $[0, N-1]$, дефиниран така:

$$w_{rec}[n] = \begin{cases} 1, & 0 \leq n < N \\ 0, & \text{иначе} \end{cases}$$

Ако умножим проста синусоида по правоъгълен прозорец, се вижда на Фигура 2.4.2, че честотното представяне се отдалечава много от "истинското", което би трябвало да представлява единична делта функция в

основната честота на синусоидата. За тази цел се въвеждат по-сложни прозоречни функции като едни от най-често ползваните са тази на Хан и тази на Хеминг, представени съответно с:

$$w_{hanning}[n] = \begin{cases} 0.5 - 0.5\,cos\,\frac{2\pi n}{N}, & 0 \leq n < N \\ 0, & \text{иначе} \end{cases} \qquad w_{hamming}[n] = \begin{cases} 0.54 - 0.46\,cos\,\frac{2\pi n}{N}, & 0 \leq n < N \\ 0, & \text{иначе} \end{cases}$$

Поведението им в честотния домейн може също да се види на Фигура 2.4.2. В описаната имплементация се ползва прозорец на Хеминг.

След като фреймовете вече са периодични, на всеки от тях се прави Фурие преобразуване. За да се възползваме от факта, че сигналът е реален, се използва Бързо Фурие преобразуване[5].

За да се моделира по-реалистично възприятието на звука, трябва да се отчете феноменът, че хората възприемат дразнителите чрез сетивата си логаритмично, в частност и звука. Този факт е отбелязан в закона на Вебер-Фехнер, а именно, че големината на усещането за определено дразнение е пропорционално на логаритъма на самото дразнение. Има различни опити да се направи скала, която по-точно да отразява човешките възприятия. Една такава е Мел скалата, която е изкуствено (емпирично) създадена. Единиците, мелове, са така избрани, че разликата между всеки два съседни да се възприема като еднаква. Връзката между меловете и честотите се задава със следната формулата:

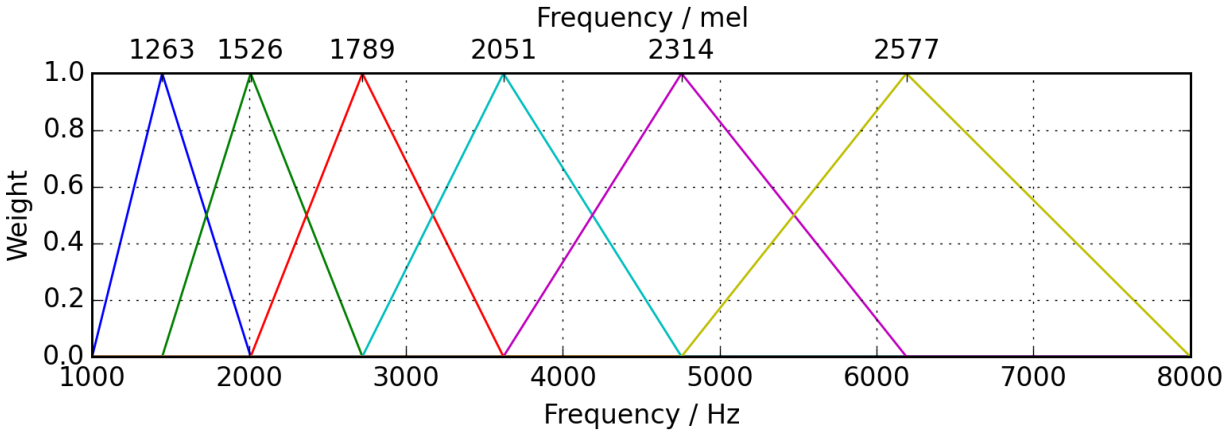
$$m = 2595\,log_{10}(1 + \frac{f}{100})$$

Втората особеност, която трябва да моделираме, е свързана с устройството на слуховия орган, тъй като целим да наподобим човешкото чуване. Главно участие има охлювчето (спирален орган), което е част от вътрешното ухо. Охлювчето е изпълнено с течност и звуковата вълна преминава през нея. По вътрешната част на охлювчето се намира така наречената базиларна мембрана, чиято дължина е покрита с рецепторни клетки - косъмчета. При преминаване на вълната през течната среда, косъмчетата се движат и предизвикват електрически сигнал, който се предава на съответните слухови неврони. Различните части на охлювчето отговарят за различни честоти (по-навътре по спиралата му са по-високите честоти). По принцип, когато два тона с различна честота стигнат до охлювчето едновременно, то ги разграничава като отделни и подава сигнал на два различни неврона. Ако тези честоти са много близки и съответно се обработват от много близки области по повърхността на охлювчето, се получава така нареченото "слухово маскиране", при което двата тона не могат да се различат като отделни и се подават на един и същи неврон. Такива области по повърхността на охлювчето се наричат критични области. При така получените Фурие коефициенти резолюцията е твърде голяма спрямо човешките възприятия и съответно отделните честоти не могат да се разграничават от охлювчето. Тъй като честотите от една критична област се възприемат като една, то

[5]на английски FFT - fast Fourier transform

можем да разделим скалата на критични области и да акумулираме информацията в тях. Това допълнително намаля пространството, в което работим, и е удобно от изчислителна гледна точка. За практически цели често се взима броят на тези области да е 23, като обикновено се взимат застъпващи и амплитудата се умножава по триъгълен прозорец. По този начин честоти, намиращи се между две съседни критични области, допринасят към акумулираните стойности и на двете области.

Като съчетаем тези две особености, получаваме застъпващи се триъгълници в Мел скалата, както е показано на Фигура 2.4.3, по които умножаваме сигнала. Взимаме логаритъм от енергията на сигнала, за да се подчертае периодичността на сигнала от глотиса, както обяснихме в предния раздел. В крайна сметка, получаваме за всеки фрейм по 23 коефициента, всеки от които стои пред акумулираните логаритми от енергии на честоти в дадена критична област.



Фигура 2.4.3: Мел скала за $16kHz$ и 6 критични области

Следващата стъпка е теорията за MFCC коефициенти, която представихме в предния раздел, да влезе в сила . За тази цел, разглеждаме новополучените 23 числа като сигнал. Правим Фурие трансформация, като запазваме информация само за реалната част (тоест косинуса), тъй като имагинерната част носи ненужната информация за фазата, до получаването на кепстър. Фундаменталната честота на глотиса и хармоничните ѝ ще образуват сигнал, чиято периодичност е засилена от логаритъма. Този „сигнал" е с голяма честота в сравнение със „сигнала", идващ от коефициентите, съответстващи на конфигурацията на вокалния тракт. Затова MFCC коефициентите, отговарящи за него, ще са в по-високите „честоти" на кепстъра. Това означава, че не са ни нужни всички MFCC коефициенти, а само тези пред ниските „честоти" на кепстъра. В имплементацията са взети първите 13 MFCC коефициента.

Последователността от горните стъпки може да се резюмира така:

- Започваме от дискретен сигнал:

  $s[n], n = 0, 1, \ldots, N$

- Разделяме сигнала $s[n]$ на феймове:

$s_t[i] = s[tS + i]$,

където $t = 0, 1, ..., \lfloor \frac{N-L}{S} \rfloor$ е t-ият фрейм, L е дължината на фрейма, а S е разстоянието между два съседни фрейма в брой дискрети.

- Прилагаме прозоречна функция:

  $x_t[n] = s_t[n] w_{hamming}[n]$, където

  $$w_{hamming}[n] = \begin{cases} 0.54 - 0.46\, cos\, \frac{2\pi n}{L}, & 0 \leq n < L \\ 0, & \text{иначе} \end{cases}$$

- Намираме Фурие коефициентите:

  $$a_{k,t} = \frac{1}{L} \sum_{n=0}^{L-1} x_t[n] e^{-\frac{2\pi i k n}{L}}$$

- Взимаме логаритъм от енергиите в критичните области:

  $c_{m,t} = log\left(\sum_{k=0}^{L-1} |a_{k,t}|^2 H_m[k, f[m-1], f[m], f[m+1]] \right), m = 0, 1, ..., M-1, M$-брой критични области.

  $$H_m[k, start, center, end] = \begin{cases} \frac{k - start}{center - start}, & start \leq k \leq center \\ \frac{end - k}{end - center}, & center < k \leq end \end{cases}$$

  $$f[m] = \frac{L}{F_s} melToHerz\left(\frac{m \times maxMel}{M+1} \right), maxMel = herzToMel\left(\frac{F_s}{2} \right),$$

  където F_s е честотата на семплиране на първоначалния сигнал.

- Правим обратно преобразуване:

  $$mfcc_t[n] = \sum_{m=0}^{M-1} c_{m,t} cos(n\pi \frac{m + 1/2}{M}), n = 0, 1, ..., M-1$$

Изменението на MFCC коефициентите във времето може да донесе допълнителна информация за вокалния тракт и подлежащата емоция. Затова в допълнение на 13-те коефициента, се добавят и първите и вторите им производни по времето, а в конкретния дискретен случай - крайни разлики. Крайният ефект е, че от входния сигнал получаваме за всеки фрейм 39 коефициента, които ще целим да класифицираме.

2.5 Класификация

След като сме избрали характеристичните вектори, които ще извличаме по подаден wav файл, трябва да можем да ги класифицираме по някакъв начин. В проучването [EKK] са разгледани и сравнени различни методи за класификация. Макар че невронните мрежи са по-често използвани в последните публикации в областта, тук ще се подходи по „старомодния" начин с Гаусови смески. Показаното за тях съотношение между „прецизност на разпознаване" и „време за трениране" е най-добро, според проучването.

Целим да намерим разпределение за всяка от търсените емоции. Всяко непрекъснато разпределение върху $\mathbb{R}^n$ може да се приближи с произволна точност с положителна линейна комбинация на достатъчно на брой гаусиани, където теглата се сумират до 1. Такава сума ще наричаме Гаусова смеска.

Нека за всяка емоция e сме приближили разпределението на векторите ѝ със смеска от K на брой гаусиани. Нека означим тази смеска с (π^e, μ^e, Σ^e), където:

$\pi^e = \{\pi_k^e\}_{k=1}^K$, тегла
$\mu^e = \{\mu_k^e\}_{k=1}^K$
$\Sigma^e = \{\Sigma_k^e\}_{k=1}^K$

Тогава при подаден нов характеристичен вектор x, ще търсим смеската на коя емоция ще доведе до най-голяма вероятностна плътност (до най-голямо правдоподобие) - тоест параметрите на кой модел е най-вероятно да са генерирали наблюдението. При подадени (π^e, μ^e, Σ^e) за дадена емоция e и характеристичен вектор x, вероятностната плътност на смеската се пресмята с формулата:

$$p(x) = \sum_{k=1}^{K} \pi_k^e \mathcal{N}(x; \mu_k^e, \Sigma_k^e),$$

където $\mathcal{N}(x, \mu, \Sigma) = \frac{exp\left(-\frac{1}{2}(x-\mu)^T \Sigma_j^{-1}(x-\mu)\right)}{\sqrt{(2\pi)^m |\Sigma|}}$ е плътност на нормално разпределение със средно μ и ковариационна матрица Σ, $\mu_i^e \in \mathbb{R}^m$, а $\sum_{k=1}^{K} \pi_k^e = 1; 0 \leq \pi_i^e \leq 1$ и в случая на избраните характеристични вектори, $m = 39$.

По принцип ковариационната матрица $\Sigma_i^e \in (\mathbb{R}^{m \times m})$, но ако приемем, че отделните MFCC коефициенти са независими, то Σ_i^e ще бъде диагонална. За да се намалят параметрите на модела, ще приемем, че това е така.

Алгоритъмът за получаване на въпросните π^e, μ^e, Σ^e е следният:

Нека $X = (x_1, ..., x_n)$ са всички характеристични вектори с етикет e. Искаме правдоподобието на тези вектори, спрямо (π^e, μ^e, Σ^e) да е възможно най-голямо. Тоест искаме да оптимизираме:

$$p(X|(\pi^e, \mu^e, \Sigma^e)) = \prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k^e \mathcal{N}(x_i; \mu_k^e, \Sigma_k^e).$$

Тъй като логаритъмът е монотонно растяща функция, то няма значение дали ще оптимизираме функцията със или без логаритъм. За по-голямо удобство, нека разглеждаме $log(p(X|(\pi^e, \mu^e, \Sigma^e)))$. Тоест имаме:

$$log(p(X|(\pi^e, \mu^e, \Sigma^e))) = log(\prod_{i=1}^{n} \sum_{k=1}^{K} \pi_k^e \mathcal{N}(x_i; \mu_k^e, \Sigma_k^e)) = \sum_{i=1}^{n} log(\sum_{k=1}^{K} \pi_k^e \mathcal{N}(x_i; \mu_k^e, \Sigma_k^e)))$$

Нека сме фиксирали някаква емоция. За удобство ще означаваме Гаусовата ѝ смеска с (π, μ, Σ). Оптимизационната задача трябва да отчита ограниченията за π, затова след добавяне на множител на Лагранж има вида:

$$L(\pi, \mu, \Sigma) = \sum_{i=1}^{n} log(\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i; \mu_k, \Sigma_k))) + \lambda(\sum_{k=1}^{K} \pi_k - 1)$$

За да максимизираме правдоподобието, търсим решения на $\frac{\partial L(\pi, \mu, \Sigma)}{\partial \mu_j} = 0, \frac{\partial L(\pi, \mu, \Sigma)}{\partial \Sigma_j} = 0$ и $\frac{L(\pi, \mu, \Sigma)}{\partial \pi_j} = 0$. Нека означим решенията на тази система съответно с $\mu_j^{new}, \Sigma_j^{new}, \pi_j^{new}$. В Приложение Г е показано, че

$$\mu_j^{new} = \frac{\sum_{i=1}^{N} \gamma_{ij} x_i}{\sum_{i=1}^{N} \gamma_{ij}}$$

$$\Sigma_j^{new} = \begin{cases} \frac{\sum_{i=1}^{N} \gamma_{ij}(x_{it} - \mu_{js})^2}{\sum_{i=1}^{N} \gamma_{ij}}, & t = s \\ 0, & \text{иначе} \end{cases}$$

$$\pi_j^{new} = \frac{\sum_{i=1}^{N} \gamma_{ij}}{N},$$

където $\gamma_{ij} = \frac{\pi_j \mathcal{N}(x_i, \mu_j, \Sigma_j)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i, \mu_k, \Sigma_k)}$

Правдоподобието относно Гаусова смеска може да се максимизира с „Алгоритъмът за максимизиране на очакването"[6] [Bis06] по следния начин:

[6]Expectation Maximisation

1. Разбиваме $X = (x_1, ..., x_n)$ на K части и взимаме за първоначални стойности на μ_j, Σ_j съответно средното и вариацията на $j-$тата част, а $\pi_j := \frac{\#_j}{|X|}$, където $\#_j =$ броят на векторите в $j-$тия клъстер.

2. Пресмятаме $\gamma_{ij} = \frac{\pi_j \mathcal{N}(x_i, \mu_j, \Sigma_j)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i, \mu_k, \Sigma_k)}$ с текущите стойности на модела.
   γ_{ij} казва каква „тежест" пада върху $j-$тата Гаусиана при генерирането на i-тия вектор.

3. Пресмятаме

$$\mu_j^{new} = \frac{\sum_{i=1}^{N} \gamma_{ij} x_i}{\sum_{i=1}^{N} \gamma_{ij}} \qquad \Sigma_j^{new} = \begin{cases} \frac{\sum_{i=1}^{N} \gamma_{ij}(x_{it} - \mu_{js}^n ew)^2}{\sum_{i=1}^{N} \gamma_{ij}}, & t = s \\ 0, & \text{иначе} \end{cases}$$

$$\pi_j^{new} = \frac{\sum_{i=1}^{N} \gamma_{ij}}{N}$$

4. Пресмятаме правдоподобието $log(p(X|(\pi^e, \mu^e, \Sigma^e)))$

5. Ако разликата между предишното и новото правдоподобие е по-малка от $\varepsilon = 1.10^{-5}$, то приключваме изпълнението с изход (π, μ, Σ), иначе се връщаме на стъпка 2 с новите стойности на модела.

Важен момент е избирането на първоначалните K клъстера. Емпирично е установено, че избирането на произволно разбиване може да забави намирането на оптимален модел. Затова често като първа стъпка се прави клъстеризация с K-means и по-точно модификацията K-means++. Разликата между K-means и K-means++ е в инициализацията на първоначалните центрове на клъстерите, които в K-means се избират на произволен принцип. В модификацията произволно се избира само първият център, а всеки от останалите $K-1$ центъра се избира вероятностно, като колкото по-отдалечена е една точка от вече избраните центрове, толкова по-вероятно е да бъде избрана. В [AV07] е показано, че намереното чрез K-means++ решение очаквано е най-много с фактор логаритъм по-лошо от оптималното.

2.6 Данни и резултати

Изборът на (свободни) емоционални бази данни за реч всъщност е доста богат. Проблемът произтича от това, че в областта рядко се прави опит за повтаряне на резултати и дори сравняване с чужди такива. Това прави поставянето на резултатите в перспектива изключително трудно.

Един от най-често използваните източници е берлинската емоционална ба-

за данни Emo-DB [Bur+05]. Изборът на специфично тази база данни е по-скоро за да може да се сравни резултатът, постигнат с гореописаните методи, отколкото заради някакво нейно преимущество (освен лесното сдобиване с нея). Emo-DB се състои от 800 записа, в които 10 актьори (5 мъже и 5 жени) изиграват 7 емоции, всяка от които е представена с 10 изречения (и няколко допълнителни втори опита). От изразените емоции избираме тези за гняв, тъга, щастие и неутрално състояние. Броят и дължините на наличните файлове са описани в Таблица 2.1

Емоция	Брой файлове	Обща дължина
Гняв	127	5 мин. 35 сек.
Щастие	71	3 мин. 00 сек.
Неутрално състояние	79	3 мин. 06 сек.
Тъга	61	4 мин. 05 сек.

Таблица 2.1: Дължина и брой файлове в emo-DB за гняв, тъга, щастие и неутрално състояние

В таблица Таблица 2.2 са показани резултати върху берлинската база данни. Изследванията използват подобни характеристики.

Източник	Тип класификатор	Тип характеристични вектори	Резултат
Текущ	GMM	MFCC	82.20%
[SCC11]	SVM	LPCMFCC	82.50%
[VA06]	Naïve Bayes classifier	MFCC	82.76%
[Gha+17]	Random forest	MFCC	79.02%

Таблица 2.2: Резултати върху emo-DB

При разглеждане на матрицата на грешките, показана в Таблица 2.3, се вижда, че класификаторът бърка емоции със сходна енергия, макар те да имат различна валентност. Това е добре известен проблем при разпознаване на емоции от реч. Точно поради тази причина речевите данни често се съчетават с допълнителен източник като видео записи, например.

	Гняв	Щастие	Неутрално	Тъга
Гняв	**91.67%**	7.50%	0.00%	0.01%
Щастие	37.14%	**54.29%**	7.14%	1.43%
Неутрално	0.00%	1.43%	**82.86%**	15.71%
Тъга	0.00%	0.00%	0.00%	**100.00%**
Общо				**82.20%**

Таблица 2.3: Матрица на грешките за база данни emo-DB на ниво файл

Втората база данни, която ще разглеждаме, е за български. Тя е компилирана (мъчително) за целите на тази дипломна работа. Състои се от записи от предаването "Тази сутрин" [BTV], които са ръчно класифицирани в четирите емоционални категории. Използвани са записи от Януари 2019 година назад до първото налично видео на сайта. Размерите и броят файлове са изложени в Таблица 2.4

Емоция	Брой файлове	Обща дължина
Гняв	51	4 мин. 15 сек.
Щастие	33	2 мин. 45 сек.
Неутрално състояние	18	1 мин. 30 сек.
Тъга	22	1 мин. 59 сек.

Таблица 2.4: Дължина и брой файлове в базата данни от "Тази сутрин" за гняв, тъга, щастие и неутрално състояние

Матрицата на грешките е показана на Таблица 2.5. Всъщност тази база данни е по-представителна за целите на дипломната работа, тъй като съдържа спонтанна реч.

За съжаление е трудно да се сравнят берлинската и местната база данни. Първо, езикът и етническата принадлежност може би играят голяма роля в изразяването на емоцията. Второ, emo-DB е записана в контролирана среда от професионални актьори, докато записите от "Тази сутрин" са хаотично записани, а професионалните говорители са умишлено избягвани. При тренирane върху берлинската бази данни и тестване с "Тази сутрин" се получава разпознаване от 39.69%, а резултат от 34.64% се наблюдава при обратната конфигурация.

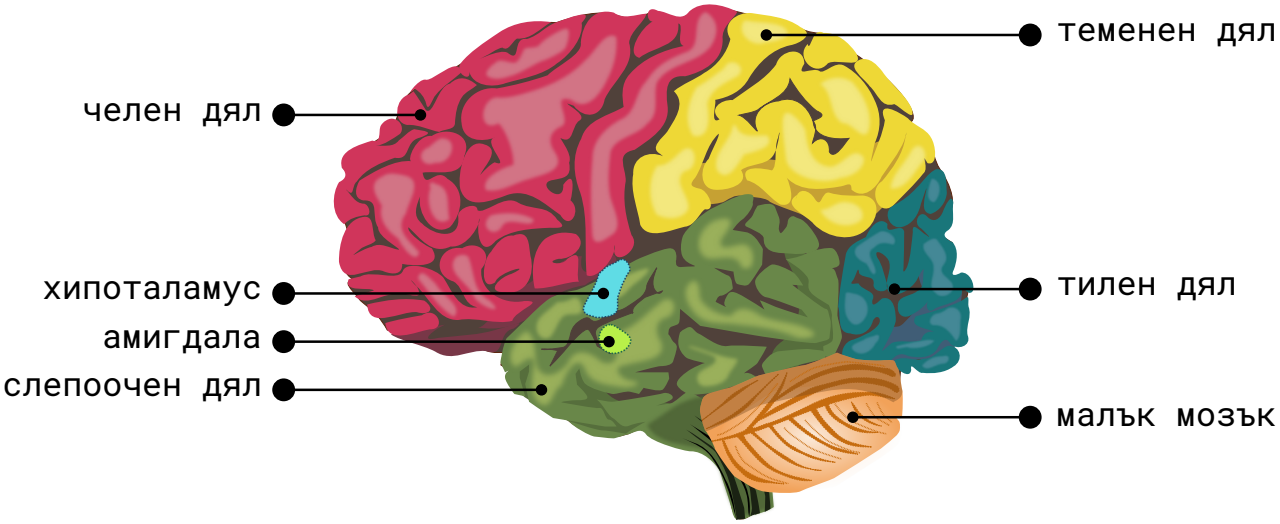
	Гняв	Щастие	Неутрално	Тъга
Гняв	**72.00%**	4.00%	12.00%	12.00%
Щастие	10.00%	**73.33%**	0.00%	16.67%
Неутрално	6.67%	6.67%	**67.67%**	20.00%
Тъга	0.00%	0.00%	0.00%	**100.00%**
Общо				**78.00%**

Таблица 2.5: Матрица на грешките за база данни от "Тази сутрин" на ниво файл

Глава 3

Сигнал от електроенцефалограф (ЕЕГ)

3.1 Грубо в мозъка



Фигура 3.1.1: Дялове на мозъка

Нервната система се разделя на централна, състояща се от главен и гръбначен мозък, и периферна. Главната функция на нервната система е да контролира работата на тялото. Информацията от околната среда се събира чрез периферната система, предава се към централната, която взима решение за съответна реакция и изпраща обратно съобщение към периферната нервна система.

Невроните (нервни клетки) са базовата функционална единица на нервната система. Те са електрически възбудими и си предават информация посредством електрически сигнали, които се предават през специални връзки между тях - синапси. Електроенцефалографът измерва колебания в напрежението на тези електрически сигнали върху повърхността на скалпа чрез множество метални пластинки, наречени електроди, долепени до него от едната страна и свързани с кабел към уреда от другата. Сигналите, получени от

електроенцефалографа, се групират по (полезни)честотни ленти по следния начин:

- $(1-4Hz)\delta$ вълни

  Асоциират се с така наречения бавновълнов сън[1], тоест най-дълбоката фаза на съня.

- $(4-8Hz)\theta$ вълни

  Има два вида θ вълни. Едните се засичат в хипокампа (частта от мозъка, свързана с формирането на спомени) и произходът им не е съвсем ясен, има разнообразни изследвания с плъхове, които изследват този вид θ-ритъм. Другите се наричат корови и са свързани с фазата на оживено сънуване[2].

- $(8-12Hz)\alpha$ вълни

  Това е най-добре изучената честотна лента. Асоциират се със спокойно будно състояние със затворени очи.

- $(13-30Hz)\beta$ вълни

  β вълните се асоциират с нормално будно състояние.

- $(30-50Hz)\gamma$ вълни (ниски)

  Според проучвания, те се свързват с изострено внимание, работеща краткосрочна и дългосрочна памет.

Главното действие на централната нервна система се осъществява в мозъчната кора, която обхваща 40% процента от обема на главния мозък. Мозъчната кора се дели на няколко дяла:

- Челен дял

  Този дял се свързва с всякакви когнитивни умения. Отговорен и за моторните функции - тоест умението да движим мускулите си доброволно.

- Теменен дял

  Той е отговорен за приемането и съчетаването на сетивна информация. Главната му функция се описва с действието „диференциране на две точки" - това е възможността на мозъка да различи, че два отделни предмета, докосващи кожата, са наистина различни, а не един.

  По този начин теменният дял участва в съставни действия като разпознаване на лица и сцени.

- Тилен дял

  Тилният дял е отговорен за зрението - разпознава заобикалящата среда, детайли и цветове в нея. В тилния дял се определя „какво", „къде" и „как" вижда човек.

[1] NREM - non-rapid eye movement sleep

[2] REM - rapid eye movement sleep. Също така чудесна музикална група

- Слепоочен дял

  Слепоочният дял се състои от структури, които са важни за дългосрочната памет. В него се намира хипокампа, който е главният дял, отговарящ за спомените. Амигдалата също е част от слепоочния дял, макар че се намира по-навътре. Смята се, че тя е отговорна за формиране на емоционален отговор, като е особено обвързана с негативните емоции.

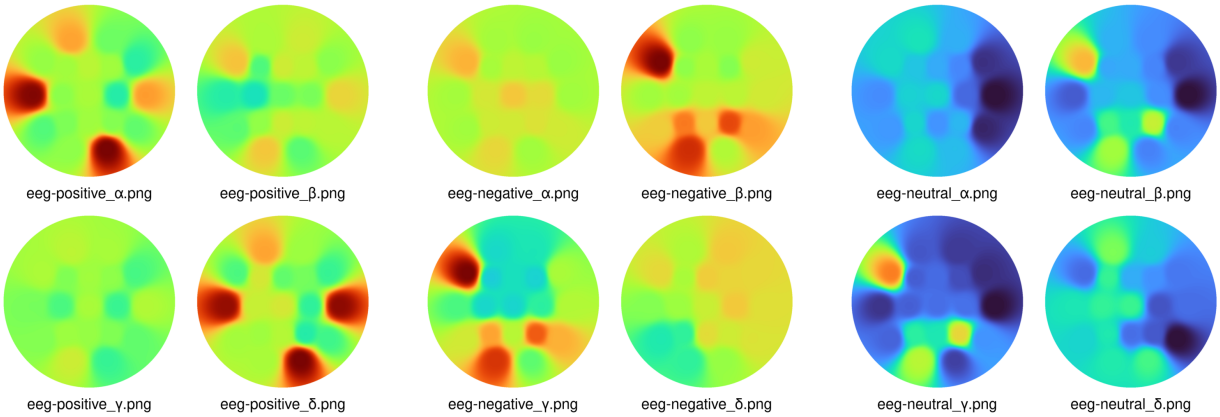
Освен на дялове, разглеждаме деленето на мозъчната кора на ляво и дясно полукълбо.

При измерването на напрежението с енцефалограф, се откриват различни модели на емоциите, в зависимост от енергията на $\theta, \beta, \alpha, \gamma$ вълните в определени дялове на мозъка.

Например по-голяма активация на α-вълни в дясната част на челния дял се свързва със стимули, които карат човек да „бяга" (от инстинкта "бий се или бягай"), тоест отговаря за негативни емоции като „погнуса" и „страх". По-голяма активност на α-вълни в ляво на челния дял се асоциира с позитивни стимули и емоции. Това означава, че асиметрията на челния дял говори за разлика във валентността. Смята се, че β и γ вълните също носят информация за валентността на емоцията. Например, при позитивни емоции γ вълните в слепоочния дял почти отсъстват, докато са с висока мощност при негативни емоции.

Активирането на амигдалата, както споменахме, също е свързано с негативни емоции. Често я наричат „зона на страха" [3].

Топлограмата от Фигура 3.1.2 показва разликата между енергиите на $\theta, \beta, \alpha, \gamma$ вълните, измерени от различните електроди, за всяка една от емоциите. Графиката е направена подобно на тези в [ZZL16]. В това изследване се цели да се намерят стабилни шаблони на емоции в ЕЕГ сигнала. Опитът се провежда върху едни и същи субекти в различен момент от време, като се разпознават четирите квадранта на активация-валентност пространството. Показва се, че активните сектори не се променят през времето.



Фигура 3.1.2: Топлограма на емоциите. Всяка една от вълните е нормализирана върху всички емоции. Червено значи висока активност, а синьо - ниска.

[3] поетично

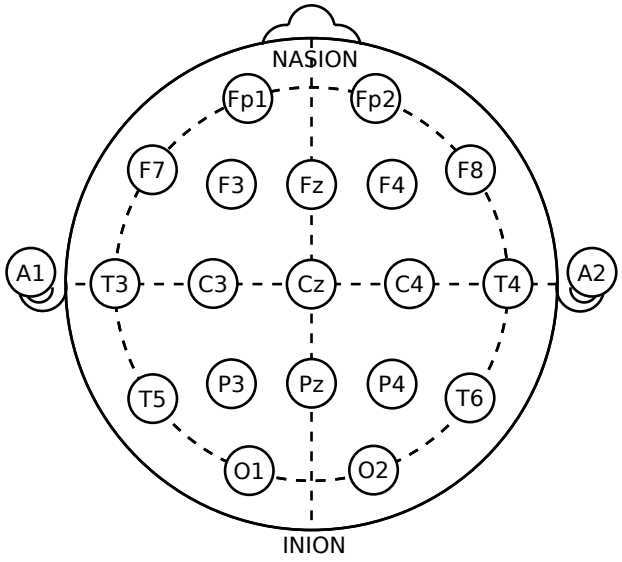
Следователно, бихме искали да изследваме енергията на $\theta, \delta, \alpha, \beta, \gamma$ вълните в различните дялове на мозъка.

3.2 Характеристики

3.2.1 Избор

За стойности на характеристичните вектори ще използваме енергията на всяка една от вълните за всеки електрод. Използването само на разликите в енергиите между двете половини на мозъка, както се предлага в много статии, не довежда до по-добър резултат върху конкретните данни. Нека разгледаме как се извличат тези характеристични вектори.

3.2.2 Извличане



Фигура 3.2.1: Система за разпределяне на електроди 10-20

Електродите се нагласят по стандартната схема 10-20. Разпределението им е показано на Фигура 3.2.1. Електроди $A1$ и $A2$ се наричат референтни. Те не са разположени по скалпа и целта им е да филтрират шума, като накрая осреднената им стойност се изважда от стойностите на останалите електроди автоматично. Електроенцефалографът записва информация за всеки от 19-те електрода с честота 500Hz, тоест за всяка секунда получаваме 500 деветнадесет-мерни вектора.

Данните от енцефалографа се прочитат, като обработката на сигнала от всеки електрод е следната:

1. Целият сигнал за даден електрод се прочита
2. Сигналът се разбива на фреймове с дължина $200ms$ и стъпка $150ms$. На всеки фрейм се прави Хеминг прозорец и се намира Фурие преобразуването
3. Енергията на Фурие коефициентите се сумира в съответните честотните ленти, отговарящи на дадена вълна. Тъй като се асоциират със състояние на дълбок сън, θ вълните не се взимат предвид. В такъв случай имаме четири честотни ленти: 4-8Hz(δ), 8-12Hz(α), 12-30Hz(β), 30-50Hz(γ)

В крайна сметка за всеки от подадените файлове, съответстващи на електроенцефалограма, получаваме за всеки електрод натрупаните стойности

в четирите честотни ленти за всеки фрейм. Ако броят на фреймовете е F, то имаме F на брой (19×4) мерни вектора.

3.3 Класификация

За класификацията се използва моделът, използващ Гаусови смески, описан в Раздел 2.5, приложен върху характеристичните вектори, описани в Раздел 3.2.

3.4 Данни и резултати

Наличните бази данни за ЕЕГ са много по-малко от тези за реч. Това вероятно се дължи на факта, че съставянето им е доста по-трудно. Освен че изискват специална апаратура (тоест електроенцефалограф), изискванията към постановката са много по-строги. Тъй като се цели при записа да има възможно най-малко дразнители, освен представените от експеримента, трябва да се подсигури, че субектът е седнал удобно на определено разстояние от монитора, не е болен, не му е студено или топло, не мига (или е със затворени очи), не говори и като цяло се движи възможно най-малко. Ако нещо в постановката се наруши, например човекът е мигнал, се трие целият запис. Често се подсигурява да не е бил консумиран кофеин (поне от страна на субекта, но вероятно е препоръчително и от страна на учения) в последните 24 часа и човекът да е имал нормален цикъл на съня. Допълнителна информация от типа дали субектът пуши, дали е левичар, или десничар и какъв е по професия, се съобщава, тъй като не се знае дали не влияе на експеримента по някакъв начин.

Целта на тази дипломна работа обаче е да се изследва комбинирането на сигнали от реч и ЕЕГ сигнали. Това изисква да имаме такава постановка, при която човекът говори, докато е свързан към електроенцефалограф, за да имаме данни от двата сигнала в един същи момент от време. Ако мигането е неприемливо, то комплексна дейност като говоренето е направо еретична за наличните бази данни. Разликата между ЕЕГ данни, извлечени докато субектът говори, и такива, при които субектът не говори, е много голяма, тъй като данните с намесена реч са много "по-шумни". Поради тази причина, и поради липсата[4] на подходяща база, такава трябваше да бъде създадена за конкретните цели.

3.4.1 Опит едно

Идеята е да се съчетаят най-често използваните похвати за създаване на емоционална речева база от една страна и емоционална ЕЕГ база данни от

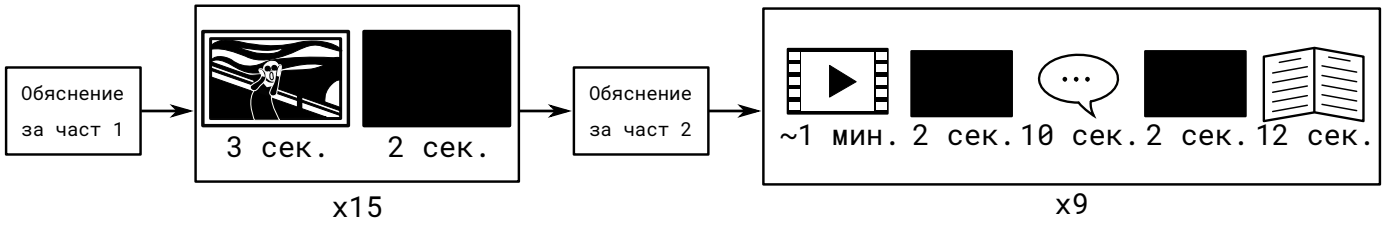
[4] до колкото ни (или поне "ми") е известно

друга.

За реч най-често се използва една от следните две постановки: субектите (най-често актьори) четат предварително избрани емоционални изречения; субектите разказват емоционална случка по свой избор, например нещо преживяно или прочетено.

За ЕЕГ най-често се използва един от следните стимули: гледане на специално подбрани емоционални картинки[5]; слушане на специално подбрана емоционална музика; гледане на специално избрани емоционални видеа.

В речевите бази данни не е нужно актьорите да изпитват емоцията, която изразяват. За сметка на това, за да се отрази в ЕЕГ сигнала, тя трябва да е неподправена. Тази задача е малко по-различна от задачата в стандартните ЕЕГ бази данни, защото тук изразяването на емоцията трябва да е по-продължително от времетраенето на картинка или видео - трябва да бъде запазена докато субектът говори.



Фигура 3.4.1

Първият опит за емоционална база данни цели резултатът да е възможно най-близък до съществуващите вече ЕЕГ бази, но все пак да се съобразява с конкретните ни изисквания. Експериментът се състои от две части. В първата част се гледат 15 специално избрани картинки (5 положителни, 5 отрицателни, 5 неутрални), като между картинките има пауза с черен екран. След това започва втората фаза, като първо на екрана се показва обяснение за протичането ѝ. Избира се едно от предварително подбраните емоционални видеа (със средна дължина от една минута), субектът описва какво е видял на видеото в продължение на десет секунди и след това чете показан на екрана текст за 12 секунди. Видеата са 3 за щастие, 4 за тъга и 3 за гняв, като са подбрани от произволни източници (YouTube, сайтове за новини и телевизионни предавания) съгласно обратната връзка на въпроса "Какво те прави тъжен/ядосан/щастлив", зададен на около петима доброволци. Текстовете, които субектът чете, са подбрани от прогнозата за времето, като се счита, че това ще произведе неутрални данни. Субектът предварително знае постановката на опита и е помолен да ограничи движенията си максимално. Постановката е показана схематично на Фигура 3.4.1, а информация за получените данни може да се види на Таблица 3.1.

След експеримента първият и единствен субект сподели, че видеата не са успели да предизвикат искрена емоция. Това означава, че информацията за емоцията ще отсъства от ЕЕГ сигнала. Оказва се, че периодът от 1-2 минути, през който се гледа видео, е твърде кратък, за да може да се събуди

[5]Например https://web.archive.org/save/https://csea.phhp.ufl.edu/media/iapsmessage.html

Емоция	Брой файлове	Обща дължина
Гняв	3	0 мин. 30 сек.
Щастие	3	0 мин. 30 сек.
Неутрално състояние	10	2 мин. 00 сек.
Тъга	4	0 мин. 40 сек.

Таблица 3.1: Дължина и брой файлове от първия опит

непринудена емоционална реакция, ако видеата не са специално избрани с предварително знание за конкретния субект. Вероятно може да се постигне желаният ефект, ако гледаните видеа са достатъчно дълги, за да може субектът да развие емоционална привързаност, например с дължината на филм. За съжаление обаче, използването на този определен вид електроди позволява максимална продължителност на целия експеримент около 10-15 минути, след което времето соленият разтвор по електродите изсъхва и съпротивлението им става твърде голямо, за да може да ги отчете уредът. Използването на картинки е безсмислено в конкретния случай, тъй като (се предполага че) събуждат емоция за части от секундата. Използването на музикални видеа изисква субектът да е чувствителен към музика.

3.4.2 Опит за разбиване на хора

Всъщност съществува проблемът, че емоциите имат твърде различен характер и заради това се предизвикват и изразяват по различен начин. Щастието се предизвиква по-лесно от гнева чрез визуални стимули. Гневът обикновено изисква субектът да има лично отношение към темата, която му се представя, независимо от стимула. Щастието има много по-момента изява от тъгата, която пък се изразява продължително, дори понякога в продължение на дни. Като цяло емоциите с ниска енергия се изразяват много по-трудно вербално от тези с висока. Но най-големият проблем идва, когато добавим и факта, че изглежда хората възприемат и изразяват емоцията по коренно различен начин един от друг.

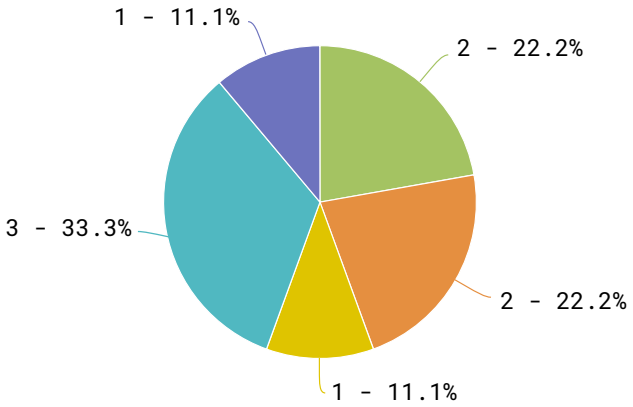
Подобен тип проблеми често се разрешава със стратегия "разделяй и владей", затова естествено възниква въпросът дали може вместо да се намерят универсални за всички хора стимули, може да се намери разбиване на хората, спрямо начинът, по който се предизвиква емоция у тях (тяхната емоционална интелигентност). Подходяща е може би класификацията на Юнг [Car21], която цели да определи типа на характера. При нея всеки човек попада в една от 16 категории, в зависимост от това къде стои по четирите оси: интровертен-екстровертен, наблюдаващ-интуитивен, чувстващ-мислещ, съдещ-приемащ[6]. Тоест дали ако двама човека попадат в една и съща категория, емоциите се предизвикват (и евентуално изразяват) по сходен начин. Класификацията на Юнг е избрана, тъй като има лесно-достъпни тестове, които "определят" класа. За целта петима участници са помолени да направят тест за класифициране на характери и след това да

[6]16 Personalities: Extraverted-Introverted, Sensing-Intuition, Thinking-Feeling,Judging-Peceiving

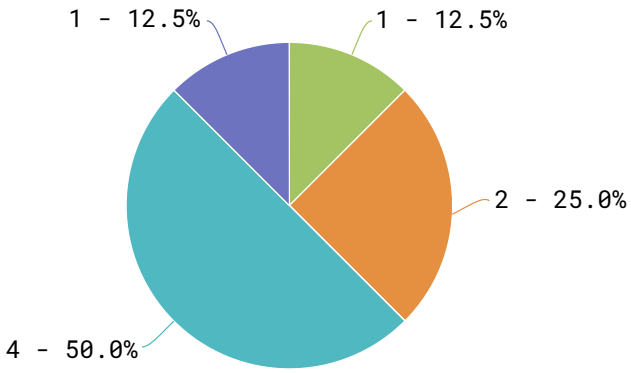
попълнят специално съставена анкета. Тя цели да провери какъв тип стимул е нужен за предизвикването на всяка една от емоциите у участниците и да покаже евентуалната връзката с класификацията на Юнг. Първо, за всяка една от изследваните емоции се търси най-подходящият стимул, като възможностите са:

1. Текст
2. Видео
3. Разговор
4. Картинка
5. Медията няма значение, само контекстът

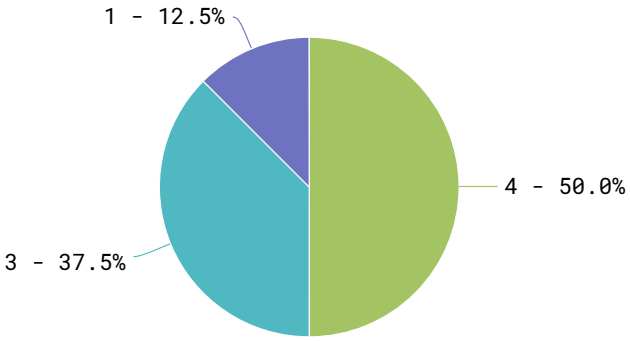
Участниците имат право да избират повече от един отговор, а резултатите са показани на фигура Фигура 3.4.2.



(а) Графика на предпочитани стимули за гняв



(б) Графика на предпочитани стимули за щастие



(в) Графика на предпочитани стимули за тъга

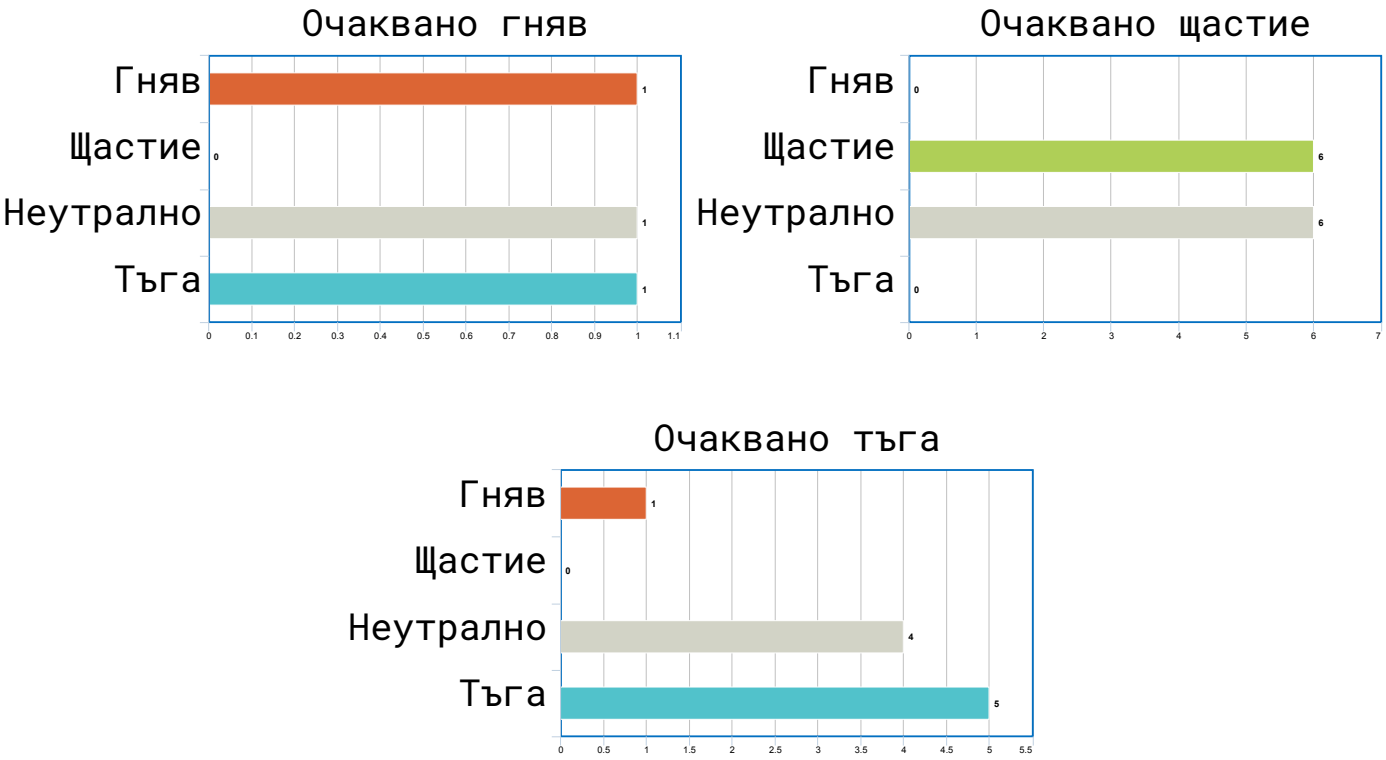
Фигура 3.4.2: Графика за предпочитаните стимули за съответните емоции

Отговорите показват, че докато тъгата може да бъде предизвикана чрез пасивни методи като текст, щастието и гневът се предизвикват по-лесно в социална обстановка - чрез разговор. За съжаление се вижда, че няма пряка връзка между класификацията на Юнг и отговорите на анкетата. Това не

означава, че по принцип този метод е неприложим, а само, че е нужна по-подходяща класификация. Повечето признати в психологията похвати изискват и професионално знание, което също излиза извън пределите на дипломната работа.

Във втората част от анкетата резултатите стават още по-объркани. В нея участниците са помолени да изберат видео, картинка, текст или да опишат нещо, което ги предразполага емоционално за всяка една от емоциите. Оказва се, че предизвикването на емоция е твърде лично и няма засичащи се теми измежду всички отговори.

Последната част на анкетата не донесе допълнителен успех. В нея на участниците са представени стимули (видеа и картинки), за които трябва да бъде определена емоцията измежду неутрално състояние, щастие, тъга и гняв. Дори тук отговорите на участниците не съвпадат за повечето емоции, както може да се види на Фигура 3.4.3.



Фигура 3.4.3: Определени емоции при представени стимули

От наличните данни не само се вижда, че хората, попадащи в един клас на Юнг, не споделят сходна чувствителност към стимулите, а че отговорите на всички участници генерално се различават. От една страна, ограничаването на възможните отговори за участниците би довело до по-голямо съвпадение на отговорите им, от друга, колкото повече трябва да променят отговорите си участниците, за да се вместят във формата, толкова по-принудена ще е и записаната емоция. Всъщност не е особено учудващо, че емоциите са толкова индивидуални. Според някои психолози ([Bar17]), емоциите изобщо не са универсални, а напротив - те са изцяло социална конструкция. Това означава, че ако участниците са от различен социален кръг (различна възраст, условия на живот, интереси), има малък шанс да имат сходен

начин за изразяването на емоциите си. Ако обстановката и опитът формират емоцията у човек, то тогава комбинациите обстановка-човек-опит са почти уникални и съответно възприемането на емоция - индивидуално. В случая нито класическата, нито алтернативната теория дават обяснение как да се предизвика непринудена емоция в повече от един участник чрез честен експеримент.

3.4.3 Опит две

Затова можем да направим по-малко честен опит, такъв, при който условията на експеримента зависят от самия субект, с цел да получим по-непринудена емоцията. С този замисъл е проведен вторият експеримент, при който емоцията се предизвиква от лични преживявания. От една страна, тъй като стимулът е избран от самия участник, не можем да гарантираме равни условия - тоест субектите могат да изберат емоционално изживяване с различна интензивност. От друга, дори постановката на опита да е еднаква за всички, това не важи и за предизвиканата емоция. В някакъв смисъл, идеята е субектите да осигурят честността на експеримента, а не самата постановка на опита.

Вторият опит се отдалечава повече от стандартните постановки за ЕЕГ бази данни, тъй като използва множество различни дразнители, целящи да предразположат максимално участниците, според собствените им нужди. Дава се възможност на субектите сами да изберат видеа, които им действат емоционално, както и аудио запис, който да слушат по време на експеримента. Останалата част протича просто: субектът си избира емоция по свое желание и сам контролира включването на стимули. При натискане на бутон “Enter” се показва видео от предварително избраните, натискане на бутон “Space” означава съответно начало и край на аудио запис, по време на който субектът е свободен да говори каквото си избере. В експеримента няма минимален брой записи, които субектът трябва да направи, нито определена тема. След първите петнайсет минути, експериментът се прекъсва (ако е продължил толкова дълго), тъй като електродите изсъхват. Субектите, участващи в експеримента, са двама и информация за получените данни е показана на Таблица 3.2

Емоция	Брой файлове	Обща дължина
Гняв	18	6 мин. 53 сек.
Щастие	0	0 мин. 00 сек.
Неутрално състояние	31	8 мин. 12 сек.
Тъга	45	15 мин. 13 сек.

Таблица 3.2: Дължина и брой файлове от втория опит

При втората постановка е намесено много движение от страна на субекта, тъй като от него се изисква да натиска бутони. Начинът, по който протича експериментът за различните субекти, може да е напълно различен. Въпреки това, резултатите върху ЕЕГ данните значително се подобряват при

данните от втория експеримент, което може да се обясни с липсата на неподправена емоция при данните от първия експеримент. Тоест, независимо, че новите данни съдържат много повече шум, те съдържат и много полезна информация.

3.4.4 Резултати

Тъй като втората постановка не съдържа положителни данни, използваме тези от първия опит, за да можем да ги сравним. Резултатите са показани съответно на Таблица 3.3 и Таблица 3.4. За пълнота, при класификацията само на речта от първия и втория експеримент се наблюдава средна класификационна точност от съответно 50.00% и 75.56%. Това показва, че данните от първия опит сдържат по-малко полезна информация и в сигнала от реч, което вероятно се дължи на факта, че субектът нито е изпитал, нито е успял да изиграе емоцията добре.

	Гняв	Щастие	Неутрално	Тъга
Гняв	**33.33%**	0.00%	33.33%	33.33%
Щастие	00.00%	**100.00%**	0.00%	0.00%
Неутрално	11.11%	0.00%	**88.89%**	0.00%
Тъга	0.00%	0.00%	100.00%	**0.00%**
Общо				**55.56%**

Таблица 3.3: Матрица на грешките от първия опит на ниво файл

	Гняв	Щастие	Неутрално	Тъга
Гняв	**100.00%**	0.00%	0.00%	0.00%
Щастие	00.00%	**100.00%**	0.00%	0.00%
Неутрално	0.00%	0.00%	**100.0%**	0.00%
Тъга	0.00%	0.00%	4.44%	**95.56%**
Общо				**98.89%**

Таблица 3.4: Матрица на грешките от втория опит на ниво файл

Тъй като данните са малко, в следващия раздел ще обединим резултата от първия и втория експеримент. Матрицата на грешките при класификацията на ЕЕГ сигнала за обединените данни може да се види на Таблица 3.5

	Гняв	Щастие	Неутрално	Тъга
Гняв	**80.33%**	5.00%	15.33%	0.00%
Щастие	25.00%	**75.00%**	0.00%	0.00%
Неутрално	0.00%	2.50%	**97.50%**	0.00%
Тъга	0.00%	2.08%	10.42%	**87.50%**
Общо				**85.00%**

Таблица 3.5: Матрица на грешките при комбиниране на данните от първия и втория опит на ниво файл

Глава 4

Двойната звезда

В предните раздели описахме получаването на класификатори за сигнали от реч и ЕЕГ поотделно. Сега въпросът е дали можем да съчетаем по някакъв начин данните или класификаторите с цел да получим по-добра класификационна точност. Разгледаните подходи са два. Първият е директен и съчетава самите характеристични вектори, а вторият - вече получените класификатори.

4.1 Съчетаване чрез конкатенация на характеристичните вектори

4.1.1 Описание

Тук идеята е възможно най-проста - конкатенираме характеристичните вектори от речта и тези от ЕЕГ сигнала до получаване на нов вектор и трениране класификатора, описан в Раздел 2.5. Единствената трудност е, че векторите за реч се взимат много по-често. За да има еднакъв брой вектори за двата сигнала, тези за речта се осредняват на всеки 200 ms.

4.1.2 Резултати

На Таблица 4.1 и Таблица 4.2 са показани средната класификационна точност за всяка една от емоциите на двата класификатора поотделно и накрая на този, получен при конкатенацията на векторите, съответно на ниво файл и на ниво вектор. Всяка от Гаусовите смески на класификатора на реч има 8 гаусиани, този за ЕЕГ е с 3 гаусиани, а полученият чрез конкатенация - 5 гаусиани - тъй като това са емпирично най-добрите стойности. От таблицата се вижда, че този метод не довежда до подобрение. Това вероятно се дължи на малкото количество данни, тъй като векторите, с които работим, са $39 + 76 = 115$ мерни. Това означава, че за да видим повече проявления на характеристиките, ще са нужни много повече данни. Освен набавянето

на допълнителни данни, като задача за бъдещо развитие може да се намали пространството чрез факторен анализ, за да се подобри обучението на класификатора.

Емоция	Само реч	Само ЕЕГ	Конкатенация
Гняв	85.00%	80.00%	86.50%
Щастие	75.00%	75.00%	57.50%
Неутрално	7.50%	97.50%	91.75%
Тъга	85.42%	87.50%	81.04%
Общо	**63.23%**	**85.00%**	**79.20%**

Таблица 4.1: Класификационна точност на класификатор за реч, класификатор за ЕЕГ и класификатор, получен при конкатенация на характеристичните вектори на ниво файл

Емоция	Само реч	Само ЕЕГ	Конкатенация
Гняв	41.53%	53.05%	68.07%
Щастие	37.35%	44.44%	34.55%
Неутрално	32.91%	66.72%	68.58%
Тъга	39.66%	70.38%	68.04%
Общо	**37.86%**	**58.65%**	**59.81%**

Таблица 4.2: Класификационна точност на класификатор за реч, класификатор за ЕЕГ и класификатор, получен при конкатенация на характеристичните вектори на ниво вектор

4.2 Съчетаване чрез максимизиране на ентропията

4.2.1 Описание

Другият похват, който е приложен, използва модел, максимизиращ ентропията. Идеята е да намерим такова разпределение p, което се държи като емпиричното разпределение върху тренировъчните данни, но в същото време не прави допълнителни предположения извън тях. Тоест имайки входни данни от вида $\mathcal{D} = (x_1, y_1), \ldots, (x_n, y_n)$, където $x_i \in X = \mathbb{R}^n$ са характеристични вектори с етикети $y_i \in Y$, където $Y = \{1, \ldots, K\}$ представя множеството от търсените емоции, искаме да максимизираме ентропията:

$$H_p(X, Y) = -\int_{x \in X} \sum_{y \in Y} p(x, y)\, log(p(x, y))$$

Нека с h_1 бележим класификатора на реч, а с h_2 този на ЕЕГ. Тогава h_1, h_2 играят роля на характеристични функции, тъй като имат вида $h_i : X \times Y \to [0, 1]$.

Очакването на всеки от класификаторите спрямо търсеното разпределение трябва да съвпада с това на емпиричното. В Приложение Д е показано,

че търсеното $\hat{p}$ има вида $\hat{p}(y|x) = \pi exp\left(\lambda_1 h_1(x,y) + \lambda_2 h_2(x,y)\right)$, където π е нормализиращ фактор и има вида $\pi = 1/\sum_{y' \in Y} exp\left(\lambda_1 h_1(x,y') + \lambda_2 h_2(x,y')\right)$. Тоест:

$$\hat{p}(y|x) = \frac{exp\left(\lambda_1 h_1(x,y) + \lambda_2 h_2(x,y)\right)}{\sum_{y' \in Y} exp\left(\lambda_1 h_1(x,y') + \lambda_2 h_2(x,y')\right)}$$

В същото приложение е показано, че е достатъчно да максимизираме логаритъм от условното правдоподобие над $\mathcal{D}$, зададено с:

$$log\left(\widehat{L}_{\mathcal{D}}(Y|X)\right) = \sum_{(x,y) \in X \times Y} \#(x,y) p(y|x)$$

С $\#(x,y)$ бележим броя на срещанията на (x,y) в корпуса.

Тоест оптимизационната задача е:

$$\begin{aligned}
\hat{\lambda}_1, \hat{\lambda}_2 &= argmax_{\lambda_1,\lambda_2} \sum_{(x,y) \in X \times Y} \#(x,y) log(p(y|x)) \\
&= argmax_{\lambda_1,\lambda_2} \sum_{(x,y) \in \mathcal{D}} log\left(\frac{exp\left(\lambda_1 h_1(x,y) + \lambda_2 h_2(x,y)\right)}{\sum_{y' \in Y} exp\left(\lambda_1 h_1(x,y') + \lambda_2 h_2(x,y')\right)}\right)
\end{aligned}$$

Намираме производните:

$$\begin{aligned}
&\frac{\partial\left[\sum_{(x,y) \in \mathcal{D}} log\left(\frac{exp\left(\lambda_1 h_1(x,y) + \lambda_2 h_2(x,y)\right)}{\sum_{y' \in Y} exp\left(\lambda_1 h_1(x,y') + \lambda_2 h_2(x,y')\right)}\right)\right]}{\partial \lambda_1} \\
&= \frac{\partial\left[\sum_{(x,y) \in \mathcal{D}} \lambda_1 h_1(x,y) + \lambda_2 h_2(x,y) - \sum_{(x,y) \in \mathcal{D}} log\left(\sum_{y' \in Y} exp\left(\lambda_1 h_1(x,y') + \lambda_2 h_2(x,y')\right)\right)\right]}{\partial \lambda_1} \\
&= \sum_{(x,y) \in \mathcal{D}} h_1(x,y) - \sum_{(x,y) \in \mathcal{D}} \frac{\sum_{y' \in Y} exp\left(\lambda_1 h_1(x,y') + \lambda_2 h_2(x,y')\right) h_1(x,y')}{\sum_{y' \in Y} exp\left(\lambda_1 h_1(x,y') + \lambda_2 h_2(x,y')\right)}
\end{aligned}$$

Съответно:

$$\frac{\partial \Lambda(\lambda_1, \lambda_2)}{\partial \lambda_2} = \sum_{(x,y) \in \mathcal{D}} h_2(x,y) - \sum_{(x,y) \in \mathcal{D}} \frac{\sum_{y' \in Y} exp\left(\lambda_1 h_1(x,y') + \lambda_2 h_2(x,y')\right) h_2(x,y')}{\sum_{y' \in Y} exp\left(\lambda_1 h_1(x,y') + \lambda_2 h_2(x,y')\right)}$$

В имплементацията λ_1 и λ_2 се получават итеративно чрез спускане по градиента с константна скорост на учене C (в имплементацията на дипломната работа $C = 1 \times 10^{-4}$).

1. Избираме $\lambda_1^0 = \lambda_2^0 = 0.5$ и пресмятаме $\hat{L}_{\mathcal{D}}^0(Y|X)$
2. За всяка стъпка t се прави следното:
   (а) Намираме $\hat{\lambda}_1$ и $\hat{\lambda}_2$
   (б) Вървим по градиента:

$$\lambda_1^t = \lambda_1^{t-1} + C\hat{\lambda}_1$$
$$\lambda_2^t = \lambda_2^{t-1} + C\hat{\lambda}_2$$

   (в) Пресмята се $\hat{L}_{\mathcal{D}}^t(Y|X)$. Ако $\hat{L}_{\mathcal{D}}^t(Y|X) \leq \hat{L}_{\mathcal{D}}^{t-1}(Y|X)$ (или $t > 200$), алгоритъмът приключва с отговор $\lambda_1^{t-1}, \lambda_2^{t-1}$

При получените по горния начин λ_1 и λ_2 и подаден вектор $x \in X$, новополученият класификатор работи по следния начин:

$$H(x) = argmax_{y \in Y} \left(\lambda_1 h_1(x, y) + \lambda_2 h_2(x, y)\right)$$

При резултата на ниво файлове няма подобрение спрямо по-добрия класификатор, както е показано в Таблица 4.3.

Емоция	Само реч	Само ЕЕГ	Комбинация
Гняв	85.00%	80.00%	80.00%
Щастие	75.00%	75.00%	75.00%
Неутрално	7.50%	97.50%	97.50%
Тъга	85.42%	87.50%	87.50%
Общо	**63.23%**	**85.00%**	**85.00%**

Таблица 4.3: Класификационна точност на класификатор за реч, класификатор за ЕЕГ и класификатор, получен при комбиниране с тегла λ_1, λ_2 на ниво файл

Ако се разгледа класификацията на отделните вектори, се вижда, че има малка разлика в полза на комбинацията (Таблица 4.4).

Емоция	Само реч	Само ЕЕГ	Комбинация
Гняв	41.53%	53.05%	53.30%
Щастие	37.35%	44.44%	44.09%
Неутрално	32.91%	66.72%	66.46%
Тъга	39.66%	70.38%	71.07%
Общо	**37.86%**	**58.65%**	**58.73%**

Таблица 4.4: Класификационна точност на класификатор за реч, класификатор за ЕЕГ и класификатор, получен при комбиниране с тегла λ_1, λ_2 ниво вектор

Големият портрет

Ако разглобиш котка, за да видиш как работи, първото нещо, което ще имаш в ръцете си, е неработеща котка[1]. Затова нека разглобим текста на дипломната работа, за да можем да направим някакво заключение.

Първо бе дадена **някаква** (неформална) дефиниция за емоция, като израз на нашето безсилие при дефиницията на понятието, и бяха избрани четири основни емоции, които да разпознаваме - щастие, гняв, тъга и неутрална емоция.

След това беше разгледана обработката на двата сигнала поотделно. Дадена бе обосновка как извличаме характеристики от речевия сигнал и защо предполагаме, че те носят информация за емоцията, съдържаща се в него. Показани са резултати от класификацията с описания в Раздел 2.5 класификатор върху няколко бази данни.

Следващата глава се занимава с обработката на ЕЕГ сигнала. След преглед как извличаме информация от електроенцефалографа и какви характеристики ще мерим в получения сигнал, текстът разглежда създаването на две бази данни за ЕЕГ сигнал. Оказа се, че тази част на дипломната работа всъщност е най-времеотнемаща, а резултатната база данни е с малък обем. И тук са показани класификационни резултати със същия класификатор.

Последната част от дипломната работа се занимава със съчетаването на двата сигнала. Разгледани са два метода. Първият, конкатенация на характеристичните вектори, води до по-ниски резултати спрямо по-добрия класификатор и е даден по-скоро за пълнота. Подобряване на този метод може евентуално да се получи след намаляне на пространството. Вторият метод на съчетаване използва тегла за двата класификатора, намерени чрез модел, максимизиращ ентропията. От резултатите се вижда, че няма допълнително подобрение на ниво файл и има минимално такова на ниво вектор.

Заключението, което можем да направим, е, че с тази постановка съчетаването на двата сигнала не довежда до подобрение.

Оттук-насетне могат да се изкажат спекулации защо.
От една страна, резултатът може да се дължи на качеството на базата данни. Тя не е създадена чрез експертизата на психолог и, поради трудностите при работа с енцефалографа, е много малка.
От друга страна, може би просто е факт, че сигналът от реч не съдържа допълнително информация, спрямо този от енцефалографа. Всъщност, в [ACC19]

[1] Последен цитат от Дъглас Адамс

показват как може да се синтезира реч директно от мозъчните вълни. Има няколко нива, на които може да се извлече информацията - първо, може да се хване "командата", която се изпраща на мускулите на вокалния тракт. Второ, имайки достъп до всеки индивидуален неврон, може да се извлече **намерението** да се изпрати тази команда. Тогава при достатъчно точно измерване (което означава **много** повече от 19 електрода), няма как сигналът за реч да носи допълнително информация, тъй като всяко действие на тялото така или иначе идва от мозъка. За съжаление, имайки стандартен електроенцефалограф, не можем да твърдим, че случаят е такъв.

И двете области - извличане на информация от реч и извличане на информация от ЕЕГ сигнал - са изключително обширни, което значи, че темата на тази дипломна работа търпи допълнително развитие. Както достатъчно пъти ми казаха обаче, човек все някъде трябва да сложи точка.

Сбогом и благодаря за рибата![2]

[2]Излъгах за последния цитат

Приложение А

Фурие приложение

А.1 Дефиниция

Понякога е по-лесно да се моделира поведението на система, ако можем да кажем как ще се държи системата за всяка честота поотделно. Например по този начин можем да нулираме всички честоти под или над дадена или да усилим определени честоти. За тази цел ни трябва еквивалентно представяне на даден сигнал във времето като съвкупност от синусоиди с различни честоти. Нека имаме дискретен във времето сигнал x, който е периодичен с фундаментален период T, измерен в секунди. Тоест,

$x(t) = x(t + T)$

Честотата, изразена в херци (периоди в секунда), се означава с $f_0 = \frac{1}{T}$ и означава "брой периоди в секунда". Нарича се фундаментална честота.

Честотата, изразена в радиани в секунда, се означава с $\omega_0 = f_0 2\pi = \frac{2\pi}{T}$ и се нарича фундаментална ъглова честота.

Тогава представянето, което търсим има вида:

$$x(t) = \sum_{k=-\infty}^{\infty} a_k e^{\frac{2k\pi it}{T}}, \tag{А.1.1}$$

където $e^{2k\pi it/T}$ е сигнал с честота $\frac{k}{T}$.

Представянето от Уравнение А.1.1 се нарича развиване в ред на Фурие за сигнала $x(t)$. Нека намерим вида на коефициентите a_k.

Умножаваме Уравнение А.1.1 с $e^{2n\pi it/T}$, тоест:

$$x(t)e^{-\frac{2n\pi it}{T}} = \sum_{k=-\infty}^{\infty} a_k e^{\frac{2k\pi it}{T}} e^{-\frac{2n\pi it}{T}}$$

Ако интегрираме двете страни от 0 до фундаменталния период T, получаваме

$$\int_0^T x(t)e^{-\frac{2n\pi it}{T}} = \int_0^T \sum_{k=-\infty}^{\infty} a_k e^{\frac{2k\pi it}{T}} e^{-\frac{2n\pi it}{T}}$$

$$\int_0^T x(t)e^{-\frac{2n\pi it}{T}} = \sum_{k=-\infty}^{\infty} a_k \left[\int_0^T e^{\frac{2(k-n)\pi it}{T}}\right]$$

Да разгледаме $\int_0^T e^{\frac{2(k-n)\pi it}{T}}$

$$\int_0^T e^{\frac{2(k-n)\pi it}{T}} = \int_0^T cos\left(\frac{2(k-n)\pi t}{T}\right) dt + i\int_0^T sin\left(\frac{2(k-n)\pi t}{T}\right) dt =$$

$$= \begin{cases} 1\Big|_0^T + 0, & n = k \\ 0 + 0, & \text{иначе} \end{cases}$$

$$= \begin{cases} T, & n = k \\ 0, & \text{иначе} \end{cases}$$

Което означава, че

$$a_n = \frac{1}{T}\int_0^T x(t)e^{-\frac{2n\pi it}{T}}$$

Това е вярно и за всеки друг интервал с дължина T:

$$a_n = \frac{1}{T}\int_T x(t)e^{-\frac{2n\pi it}{T}} \tag{A.1.2}$$

Може да се покаже [OWN96], че редът на Фурие за сигнал $x(t)$ е сходящ и съответно коефициентите от Уравнение A.1.2 са крайни, ако е изпълнено че:

$$\int_T |x(t)|^2 < \infty,$$

Още повече, ако сигналът x е дискретен и периодичен (какъвто е случаят, когато семплираме речеви сигнал) и периодичен $x[n] = x[n+N]$, имаме само N различни стойности:

$e^{\frac{2(k+N)\pi in}{N}} = e^{\frac{2k\pi in}{N}} e^{\frac{2N\pi in}{N}} = e^{\frac{2k\pi in}{N}}$, тъй като $e^{2\pi in} = cos(2\pi n) + sin(2\pi n) = 1$

следователно са ни достатъчни само кои да е N последователни стойности:

$$x[n] = \sum_{k=-\infty}^{\infty} \hat{a}_k e^{\frac{2k\pi in}{N}}$$

$$x[n] = \sum_{k=0}^{N-1} a_k e^{\frac{2k\pi in}{N}} \tag{A.1.3}$$

Уравнение A.1.3 се нарича ред на Фурие за дискретен във времето сигнал.

Коефициентите можем да намерим по същия начин като в непрекъснатия случай, но използвайки сума, вместо интеграл:

$$\sum_{n=0}^{N-1} x[n]e^{\frac{-2\pi irn}{N}} = \sum_{n=0}^{N-1}\sum_{k=0}^{N-1} a_k e^{\frac{2\pi ikn}{N}} e^{\frac{-2\pi irn}{N}} =$$
$$\sum_{n=0}^{N-1} x[n]e^{\frac{-2\pi irn}{N}} = \sum_{k=0}^{N-1} a_k \sum_{n=0}^{N-1} e^{\frac{2\pi i(k-r)n}{N}}$$

и отново използваме, че

$$\sum_{n=0}^{N-1} e^{\frac{2\pi i(k-r)n}{N}} = \begin{cases} N, & k-r \equiv 0(modN) \\ 0, & \text{иначе} \end{cases}$$
$$\Rightarrow a_r = \frac{1}{N}\sum_{n=0}^{N-1} x[n]e^{-\frac{2\pi irn}{N}} \tag{A.1.4}$$

което е изпълнено и за всеки друг интервал с дължина N. Ще използваме означенията $x(t) \overset{\mathcal{FS}}{\longleftrightarrow} a_k$ или $x(t) \overset{\mathcal{FS}}{\longleftrightarrow} X(e^{i\omega_k})$,

където $a_k = X(e^{\frac{2\pi ik}{N}}) = X(e^{i\omega_k})$ за $\omega_k = \frac{2\pi k}{N}$

A.2 Свойства

- Изместване във времето

  Ако $x[n] \overset{\mathcal{FS}}{\longleftrightarrow} a_k$, то $x[n-n_0] \overset{\mathcal{FS}}{\longleftrightarrow} b_k = a_k e^{-\frac{2\pi in_0}{N}}$

  Тъй като Уравнение A.1.4 е изпълнено за всеки интервал, то можем да изберем интервала $[n_0, T-1+n_0]$

  $$b_k = \frac{1}{N}\sum_{n=n_0}^{N-1+n_0} x[n-n_0]e^{-\frac{2\pi ikn}{N}} = \sum_{n=n_0}^{N-1+n_0} x[n-n_0]e^{-\frac{2\pi ik(n-n_0)}{N}} e^{-\frac{2\pi ikn_0}{N}} =$$
  $$e^{-\frac{2\pi ikn_0}{N}} \sum_{\tau=0}^{T} x[\tau]e^{-\frac{2\pi ik\tau}{N}} = e^{-\frac{2\pi ikn_0}{N}} a_k$$

- Симетричност на комплексно спрегнатите за реален сигнал

  Ако $x[n] = \bar{x}[n]$ е реален сигнал, за който $x(t) \overset{\mathcal{FS}}{\longleftrightarrow} a_k$, то $\overline{a}_k = a_{-k}$

От уравнение Уравнение A.1.4 следва, че:

$$a_k = \frac{1}{N}\sum_{n=0}^{N-1} x[n]e^{-\frac{2\pi ikn}{N}}$$

$$\overline{a_k} = \frac{1}{N}\sum_{n=0}^{N-1} \overline{x[n]e^{-\frac{2\pi ikn}{N}}} = \frac{1}{N}\sum_{n=0}^{N-1} \overline{x[n]}e^{\frac{2\pi ikn}{N}} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]e^{\frac{2\pi ikn}{N}}$$

$$\Rightarrow \overline{a_k} = a_{-k} = a_{N-k}$$

Това означава, че за реални сигнали са достатъчни първите $\frac{N}{2}+1$ коефициенти, тъй като останалите са им комплексно спрегнати. Честотата, която отговаря на $a_{\frac{N}{2}}$ се нарича Найкуист честота.

A.3 Конволюция

Често ще се налага да използваме връзката между умножение, Фурие трансформация и операцията конволюция.

Дефиниция. (Дискретна конволюция)
Ако $f, g : \mathbb{N} \mapsto \mathbb{Z}$, дискретна конволюция (конволюционна сума) на f и g, наричаме

$$(f * g)[n] = \sum_{k=-\infty}^{\infty} f[k]g[n-k]$$

Ако f и g са периодични с период N, то $(f * g)[n] = \sum_{k=0}^{N-1} f[k]g[n-k]$

Пример 1 (Теорема за конволюцията за периодични дискретни сигнали). Ако $f[n] \overset{\mathcal{FS}}{\longleftrightarrow} F(e^{i\omega_k})$ и $g[n] \overset{\mathcal{FS}}{\longleftrightarrow} G(e^{i\omega_k})$ и f, g са периодични с период N.

$(f*g)[n] \overset{\mathcal{FS}}{\longleftrightarrow} F(e^{i\omega_k}) \cdot G(e^{i\omega_k})$ Дуалното твърдение също е вярно за непрекъснатия вариант на конволюция.

$$F(e^{i\omega_k}) = a_k = \frac{1}{N}\sum_{n=0}^{N-1} f[n]e^{-\frac{2\pi ikn}{N}}$$

$$G(e^{i\omega_k}) = b_k = \frac{1}{N}\sum_{n=0}^{N-1} g[n]e^{-\frac{2\pi ikn}{N}}$$

Нека $h[n] = (f * g)[n]$ и $(f * g)[n] \overset{\mathcal{FS}}{\longleftrightarrow} H(e^{i\omega_k})$.

Тогава:

$$H(e^{i\omega_k}) = c_k = \frac{1}{N}\sum_{n=0}^{N-1} h[n]e^{-\frac{2\pi ikn}{N}}$$

$$= \frac{1}{N}\sum_{n=0}^{N-1}(f*g)[n]e^{-\frac{2\pi ikn}{N}} = \frac{1}{N}\sum_{n=0}^{N-1}\sum_{r=0}^{N-1} f[r]g[n-r]e^{-\frac{2\pi ikn}{N}}$$

$$= \frac{1}{N}\sum_{r=0}^{N-1} f[r]\sum_{n=0}^{N-1} g[n-r]e^{-\frac{2\pi ikn}{N}}$$

$$= \frac{1}{N}\sum_{r=0}^{N-1} f[r]\left[\sum_{n=0}^{N-1} g[n-r]e^{-\frac{2\pi ikn}{N}}\right]$$

Нека положим y=n-r

$$= \frac{1}{N} \sum_{r=0}^{N-1} f[r] \left[\sum_{y=-r}^{N-1-r} g[y] e^{-\frac{2\pi i k(y+r)}{N}} \right] = \frac{1}{N} \sum_{r=0}^{N-1} f[r] \left[\sum_{y=-r}^{N-1-r} g[y] e^{-\frac{2\pi i k y}{N}} \right] e^{-\frac{2\pi i k r}{N}}$$

Понеже $[-r, N-1-r]$ е интервал с дължина N е изпълнено:

$$= \frac{1}{N} \sum_{r=0}^{N-1} f[r] e^{-\frac{2\pi i k r}{N}} N G(e^{i\omega_k}) = N F(e^{i\omega_k}) \cdot G(e^{i\omega_k})$$

Приложение Б

Приложение за полюси и нули

Б.1 Дефиниция

Нека $z \in \mathbb{C}$. Видяхме,че предавателната функция $\mathcal{H}$ на определени системи (и в частност филтри) има вида:

$$\mathcal{H}(z) = \frac{\sum_{m=0}^{M} b_m z^{-m}}{\sum_{k=0}^{N} a_k z^{-k}} = \tag{2.3.6}$$

$$= \frac{N(z)}{D(z)} = G\frac{(z-\beta_1)(z-\beta_2)...(z-\beta_M)}{(z-\alpha_1)(z-\alpha_2)...(z-\alpha_N)},$$

където $G = b_0/a_0$ и се нарича усилващ коефициент.

С β_i означаваме корените на уравнението $N(z) = 0$. Те се наричат нули на системата и $\lim_{z\to\beta_i} \mathcal{H}(z) = 0$

С α_i означаваме корените на уравнението $Dz(z) = 0$. Те се наричат полюси на системата и $\lim_{z\to\alpha_i} \mathcal{H}(z) = \infty$

Тъй като коефициентите на $N(z)$ и $D(z)$ са реални, нулите (и съответно полюсите) ще са са или реални, или са част от двойка комплексно спрегнати. Тоест, няма нула (или полюс), която да е комплексна, но да няма комплексно спрегнато из останалите нули (полюси).

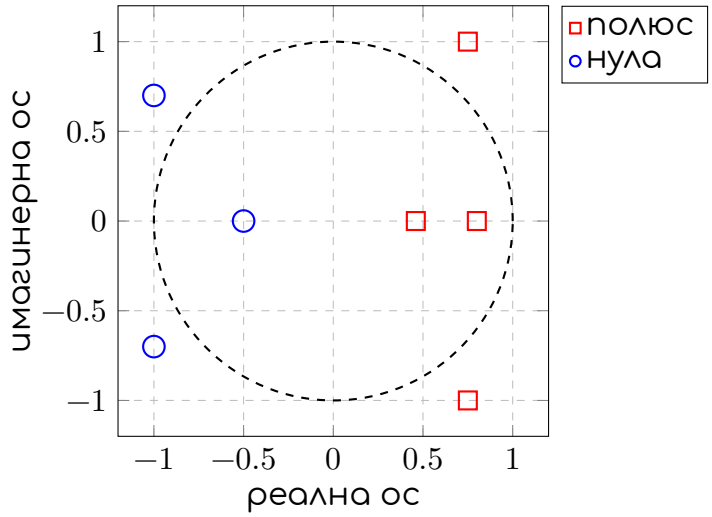
Това представяне е удобно, защото ни позволява да разбием $\mathcal{H}$ на произведение от по-прости предавателни функции:

$$\mathcal{H}(z) = \underbrace{G\frac{(z-\beta_1)}{(z-\alpha_1)}}\ \underbrace{\frac{(z-\beta_2)}{(z-\alpha_2)}}\ ...\ \underbrace{\frac{(z-\beta_M)}{(z-\alpha_N)}}$$

$$\mathcal{H}(z) = \mathcal{H}_1(z)\mathcal{H}_2(z)\,...\,\mathcal{H}_K(z),$$

където $\mathcal{H}_i$ е произведение на няколко полюса и нули.

Тоест, достатъчно е да видим какви филтри се описват от трансферни функции, съдържащи една или две нули и полюси, за да можем да направим извод за целия филтър $\mathcal{H}$.



Фигура Б.1.1: Полюс-нула графика

Фигура Б.1.1 изобразява трансферна функция с три нули и четири полюса, от които една реална нула и два реални полюса. Нулите и полюсите, които не са реални, са комплексно спрегнати.

Б.2 Характеризация на филтри

Една система се описва изцяло от трансферната си функция, а всяка трансферна функция може да се представи като произведение на нули и полюси. Следователно, анализирайки тези нули и полюси, можем да направим извод за действието на филтъра.

Нека имаме следния филтър от първи ред:

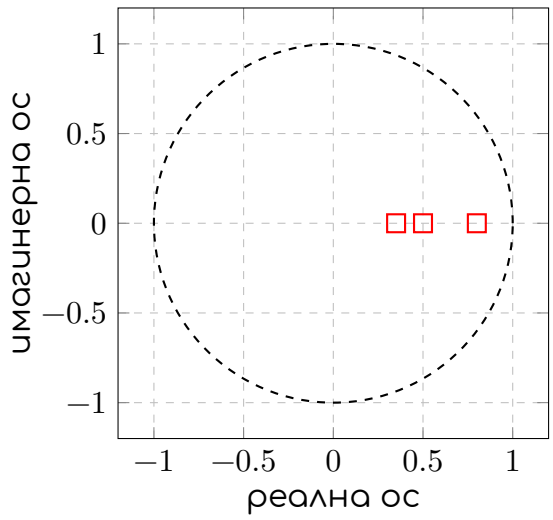
$$y[n] = b_0 x[n] + a_1 y[n-1] \overset{\mathcal{FS}}{\longleftrightarrow}$$

$$Y(z) = b_0 X(z) + a_1 z^{-1} Y(z) \longleftrightarrow$$

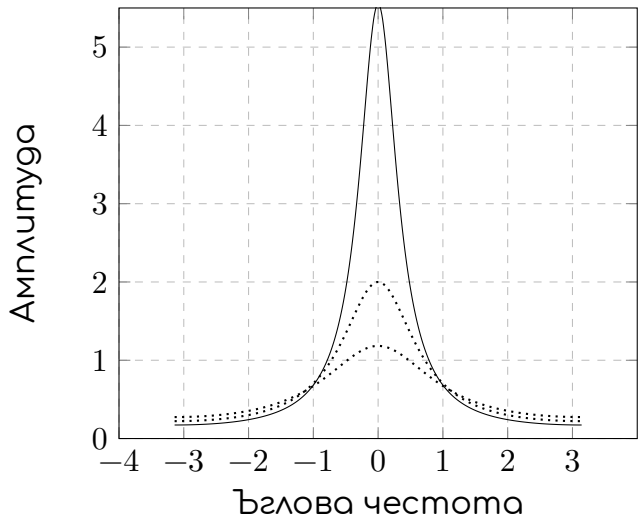
$$\frac{Y(z)}{X(z)} = \frac{b_0}{1 - a_1 z^{-1}} \longleftrightarrow$$

$$\mathcal{H}(z) = G \frac{1}{1 - a_1 z^{-1}}, G = b_0$$

Тъй като имаме само един полюс, то следва, че a_1 е реално число, тъй като няма как да е част от комплексно спрегната двойка. Това означава, че a_1 напълно описва вида на $\mathcal{H}$, а b_0 играе ролята на усилващ коефициент.

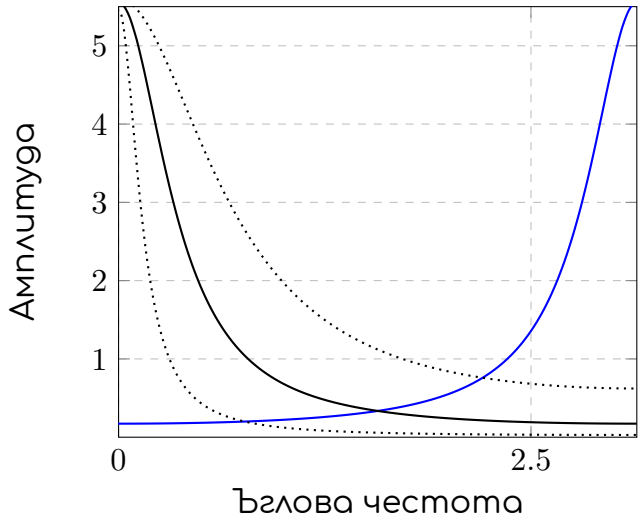


(а) Полюс-нула графика за $\mathcal{H}$



(б) Графика на $g(\omega)$. От горе надолу:$a = 0.7, a = 0.5, a = 0.35$

Фигура Б.2.1: Действие на филтър от първи ред за $a = 0.35, a = 0.5, a = 0.7$



Фигура Б.2.2: Действие на филтър от първи ред за различни стойности на a и b
С черно от ляво надясно:
a = 0.5, b = 2.8;
a = 0.7, b = 1;
a = 0.865, b = 0.2
Със синьо:
a = -0.7, b = 1

Понеже a_1 е реално число, винаги ще лежи на реалната ос, както е показано на полюс-нула графиката на Фигура Б.2.2

Нека разгледаме $\mathcal{H}$ в честотния домейн: $\mathcal{H}(e^{i\omega}) = \dfrac{b_0}{1 - a_1 e^{-i\omega}}$, където ω е ъглова честота, измерена в радиани. Можем изразим $\mathcal{H}$ като функция на ω:

$$\mathcal{H}(e^{i\omega}) = \frac{b_0}{1 - a_1 e^{-i\omega}} = \frac{b_0}{1 - a_1\, cos\,\omega + i a_1\, sin\,\omega} = \frac{b_0(1 - a_1\, cos\,\omega - i a_1\, sin\,\omega)}{1 - 2a_1\, cos\,\omega + a_1^2}$$

$$= \frac{b_0(1 - a_1\, cos\,\omega)}{1 - 2a_1\, cos\,\omega + a_1^2} + i\frac{-\, b_0 a_1\, sin\,\omega}{1 - 2a_1\, cos\,\omega + a_1^2}$$

Нека с $g(\omega)$ означим модула на $\mathcal{H}$

$$g(\omega) = \frac{b_0^2(1 - 2a_1\, cos\,\omega + a_1^2)}{(1 - 2a_1\, cos\,\omega + a_1^2)^2} = \frac{b_0^2}{1 - 2a_1\, cos\,\omega + a_1^2}$$

На Фигура Б.2.2 се вижда графиката на $g(\omega)$ за различни стойности на a_1 и b_0. Този вид филтри се наричат **резонатори**, тъй като честотите във върха на графиката ще се усилят. Резонаторите се описват главно чрез своята **амплитуда** - височината на максимума, **честота** - къде е върхът върху честотната ос, **честотна лента** - колко е широка графиката, което определя колко честоти ще се усилят.

В случая на филтър от първи ред, амплитудата и честотната лента се определят от a_1 и b_0, а върха на графиката винаги ще е в 0. Тоест този вид филтри могат да усилват само честотите около 0.

При $a_1 > 0$, филтрите се наричат **нискочестотни**, защото пропускат ниските честоти и задържат високите (с черно на Фигура Б.2.2).

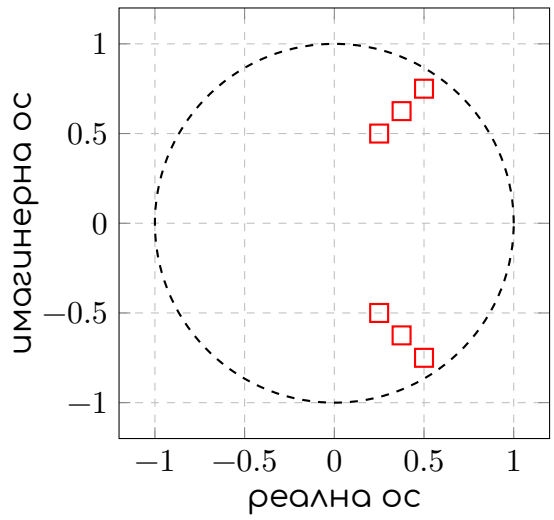
При $a_1 < 0$, филтрите се наричат **високочестотни** (в синьо на Фигура Б.2.2)

За да се премести пикът на функцията нанякъде по честотната ос извън нулата, трябва a_1 да е комплексно. Ако трансферната функция има само един полюс, a_1 винаги е реално, затова ни трябва поне една комплексно спрегната двойка. Нека разгледаме филтър от втори ред.

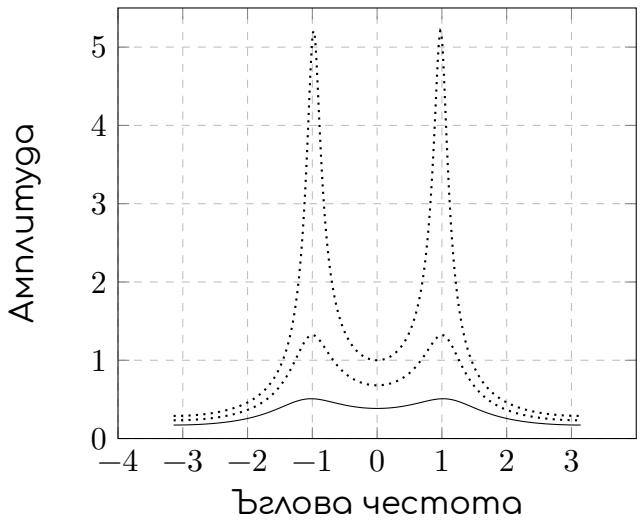
$$y[n] = b_0 x[n] + a_1 y[n-1] + a_2 y[n-2]$$

$$\mathcal{H}(z) = \frac{b_0}{1 - a_1 z^{-1} - a_2 z^{-2}}$$

$$\mathcal{H}(z) = G\frac{1}{(1 - \alpha_1 z^{-1})(1 - \alpha_2 z^{-1})}$$

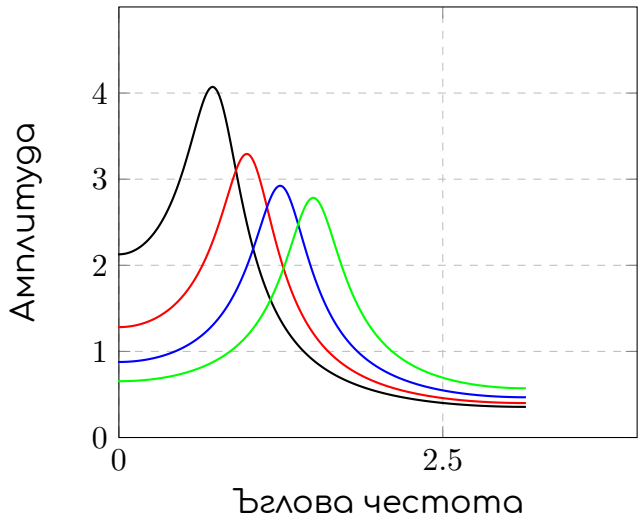


(а) Полюс-нула графика за $\mathcal{H}$

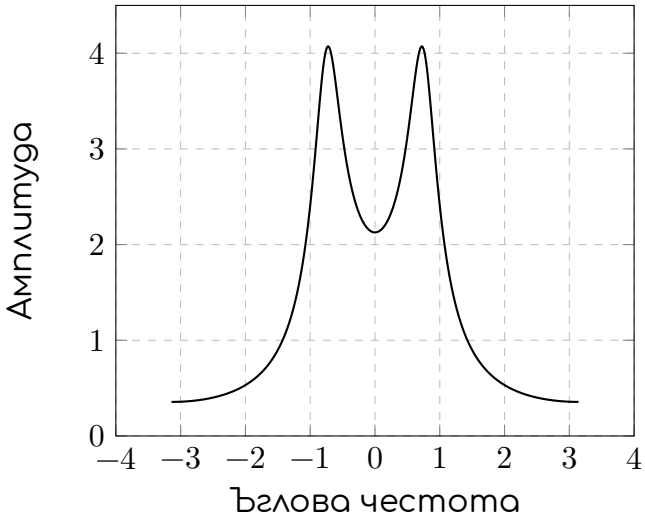


(б) Графика на $g(\omega)$ в $-\pi, \pi$

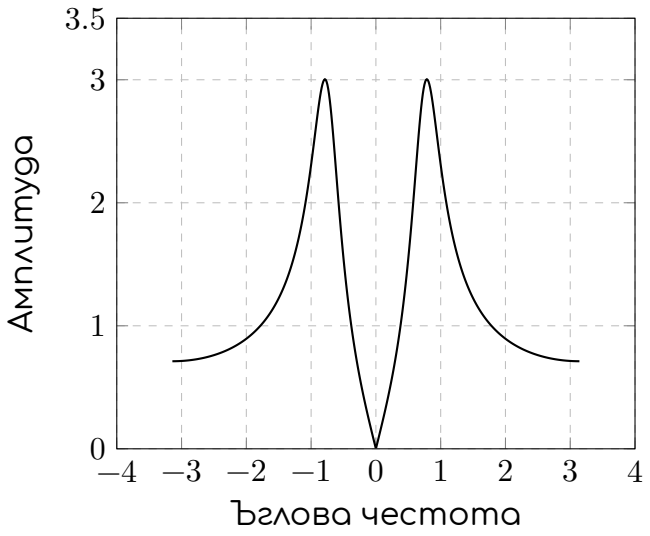
Фигура Б.2.3: Действие на филтър от втори ред за $\alpha_1 = (0.25 + 0.5i), (0.5 + 0.75i), (0.375 + 0.625)$



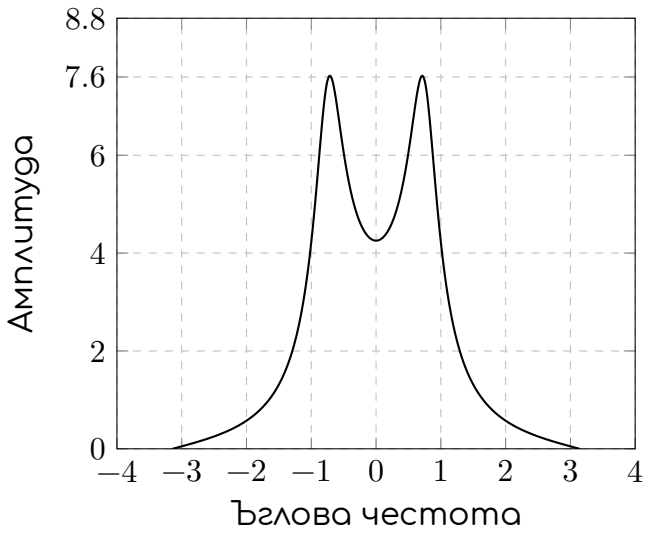
Фигура Б.2.4: Графика на g с отдалечаващи се от реалната ос полюси



(а) $b_0 = 1, b_1 = 0$



(б) $b_0 = 1, b_1 = 1$

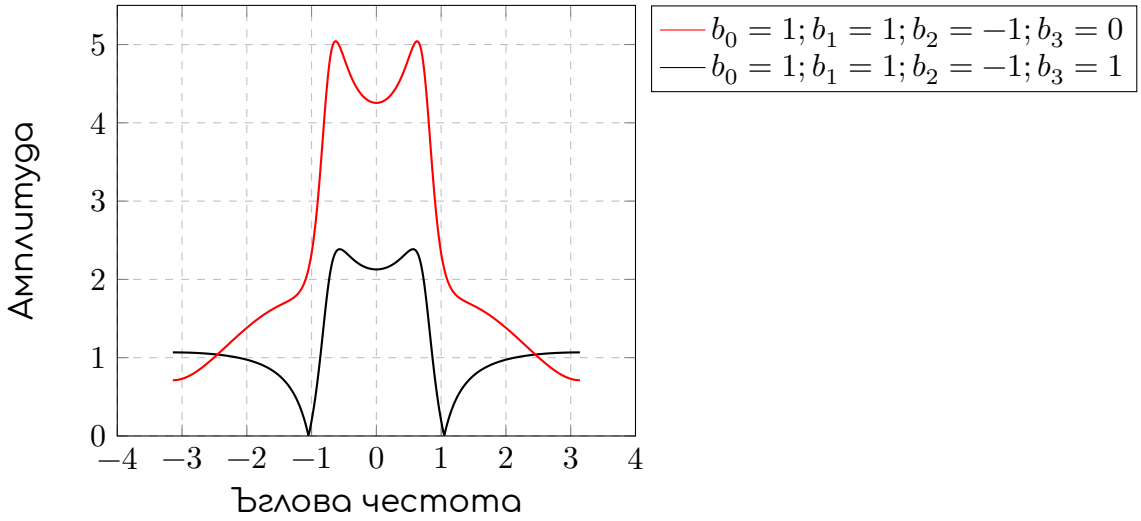


(в) $b_0 = 1, b_1 = -1$

Фигура Б.2.5: Действие на филтър от втори ред за различни стойности на b и $a_1 = 1.17, a_2 = -0.64$

Местенето на полюсите по-далеч от реалната ос, раздалечава върховете по честотната лента, както се вижда на Фигура Б.2.4

Видът на резонатора (тоест честотна лента, честота и амплитуда), се определят главно от полюсите. Добавянето на нули също влияе на вида на филтъра, както може да се види от Фигура Б.2.5 В единият случай се добавя нула в нулата, в другия - в края на спектъра.



Фигура Б.2.6: Действие на филтър от вида $\mathcal{H}(e^{i\omega}) = \frac{b_0 - b_1 e^{-i\omega} - b_2 e^{-2i\omega} - b_3 e^{-3i\omega}}{1 - a_1 e^{-i\omega} - a_2 e^{-2i\omega}}$ за $a_1 = 1.17, a_2 = -0.64$

Добавянето на допълнителни нули може да се види на Фигура Б.2.6. Тези нули се наричат **антирезонанси**.

Тогава можем да разложим даден сложен филтър $\mathcal{H}$ по следния начин:

$$\mathcal{H}(z) = \mathcal{H}_1(z)\mathcal{H}_2(z)\ldots\mathcal{H}_K(z),$$

където H_i е по-прост филтър от първи или втори ред, чийто вид може лесно да се моделира чрез промяна на коефициентите.

След това съчетаването на простите е просто произведение в честотния домейн, а свойствата на Фурие преобразуванията ни дават вида и във времевия домейн.

Приложение В

Приложение към Сигнал от реч

Пример 2. $\mathcal{V}(z)$ за $N = 2$ и произволно τ_i

Имаме:

$U_k = Q_k U_{k+1}$ за

$$U_k = \begin{bmatrix} U_k^+(z) \\ U_k^-(z) \end{bmatrix}$$

$$Q_k = \begin{bmatrix} \dfrac{z^{\tau_k}}{1+r_k} & \dfrac{-r_k z^{\tau_k}}{1+r_k} \\ \dfrac{-r_k z^{-\tau_k}}{1+r_k} & \dfrac{z^{-\tau_k}}{1+r_k} \end{bmatrix} = z^{\tau_k} \begin{bmatrix} \dfrac{1}{1+r_k} & \dfrac{-r_k}{1+r_k} \\ \dfrac{-r_k z^{-2\tau_k}}{1+r_k} & \dfrac{z^{-2\tau_k}}{1+r_k} \end{bmatrix} = z^{\tau_k} \hat{Q}_k$$

Търсим: $V(z)$

Доказателство:

$$\frac{1}{\mathcal{V}(z)} = \frac{U_G(z)}{U_L(z)} = \begin{bmatrix} \dfrac{2}{1+r_G}, & -\dfrac{2r_G}{1+r_G} \end{bmatrix} \prod_{i=1}^{N} Q_i \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \tag{2.2.23}$$

$$= z^{(\tau_1+\tau_2)} \begin{bmatrix} \frac{2}{1+r_G}, & -\frac{2r_G}{1+r_G} \end{bmatrix} \hat{Q}_1\hat{Q}_2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} =$$

$$= z^{(\tau_1+\tau_2)} \begin{bmatrix} \frac{2}{1+r_G}, & -\frac{2r_G}{1+r_G} \end{bmatrix} \begin{bmatrix} \frac{1}{1+r_1} & \frac{-r_1}{1+r_1} \\ \frac{-r_1 z^{-2\tau_1}}{1+r_1} & \frac{z^{-2\tau_1}}{1+r_1} \end{bmatrix} \begin{bmatrix} \frac{1}{1+r_2} & \frac{-r_2}{1+r_2} \\ \frac{-r_2 z^{-2\tau_2}}{1+r_2} & \frac{z^{-2\tau_2}}{1+r_2} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} =$$

$$= 2z^{(\tau_1+\tau_2)} \begin{bmatrix} \frac{1+r_G r_1 z^{-2\tau_1}}{(1+r_G)(1+r_1)}, & -\frac{r_1+r_G z^{-2\tau_1}}{(1+r_G)(1+r_1)} \end{bmatrix} \begin{bmatrix} \frac{1}{1+r_2} & \frac{-r_2}{1+r_2} \\ \frac{-r_2 z^{-2\tau_2}}{1+r_2} & \frac{z^{-2\tau_2}}{1+r_2} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} =$$

$$= 2z^{(\tau_1+\tau_2)} \begin{bmatrix} \frac{1+r_G r_1 z^{-2\tau_1}+r_1 r_2 z^{-2\tau_2}+r_G r_2 z^{-2(\tau_1+\tau_2)}}{(1+r_G)(1+r_1)(1+r_2)}, & -\frac{r_2+r_G r_1 r_2 z^{-2\tau_1}+r_1 z^{-2\tau_2}+r_G z^{-2(\tau_1+\tau_2)}}{(1+r_G)(1+r_1)(1+r_2)} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\iff$$

$$\mathcal{V}(z) = \frac{0.5z^{-(\tau_1+\tau_2)}(1+r_G)\prod_{i=1}^{2}(1+r_i)}{1+r_G r_1 z^{-2\tau_1}+r_1 r_2 z^{-2\tau_2}+r_G r_2 z^{-2(\tau_1+\tau_2)}}$$

Бележка: Тъй като r_G, r_1, r_2 са ненулеви, то за да получим максимална степен, без да имаме нулеви коефициенти, трябва:

$$\begin{vmatrix} 2\tau_1 = 1 \\ 2\tau_2 = 1 \\ 2(\tau_1+\tau_2) = 2 \end{vmatrix} \tag{B.0.1}$$

Тоест $\tau_1 = \tau_2 = \frac{1}{2}$

Пример 3. $\mathcal{V}(z)$ за $N = 2$ и произволно $\tau_1 = \tau_2 = \frac{1}{2}$

Доказателство:

Заместваме $\tau_i = \frac{1}{2}$ в Пример 2

$$\mathcal{V}(z) = \frac{0.5z^{-1}(1+r_G)\prod_{i=1}^{2}(1+r_i)}{1+(r_G r_1+r_1 r_2)z^{-1}+r_G r_2 z^{-2}}$$

Приложение Г

Приложение към Класификация

Свойство 2. Нека $x \in \mathbb{R}^m, A \in \mathbb{R}^m \times \mathbb{R}^m$, A е диагонална. Тогава $\frac{\partial x^T Ax}{\partial x} = 2Ax$

Да разгледаме производната по някоя от координатите - x_k

$$\frac{\partial x^T Ax}{\partial x_k} = \frac{\partial \left(x_1, x_2, \dots x_m\right) \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{mm} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}}{\partial x_k}$$

$$= \frac{\partial \left(x_1 a_{11}, x_2 a_{22}, \dots x_m a_{mm}\right) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}}{\partial x_k}$$

$$= \frac{\partial \left(x_1^2 a_{11} + x_2^2 a_{22} + \dots + x_m^2 a_{mm}\right)}{\partial x_k} = 2x_k a_{kk}$$

Това означава, че:

$$\frac{\partial x^T Ax}{\partial x} = (2x_1 a_{11}, 2x_2 a_{22} \dots 2x_m a_{mm}) = 2Ax$$

Свойство 3. Ако A е диагонална матрица, $A = (a_{ii})_{i=1}^m$, $\frac{\partial |A|}{\partial a_{ii}} = \frac{|A|}{a_{ii}}$

$$\frac{\partial |A|}{\partial a_{ii}} = \frac{\partial \left(\prod_{i=1}^{m} a_{ii}\right)}{\partial a_{ii}} = a_{11}.a_{22} \dots a_{i-1i-1} a_{i+1i+1} \dots a_{mm} = \frac{\prod_{i=1}^{m} a_{ii}}{a_{ii}} = \frac{|A|}{a_{ii}}$$

Нека $L(\pi, \mu, \Sigma) = \sum_{i=1}^{n} log(\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k))) + \lambda(\sum_{k=1}^{K} \pi_k - 1)$

$$\mathcal{N}(x_i, \mu_j, \Sigma_j) = \frac{exp\left(-\frac{1}{2}(x_i - \mu_j)^T \Sigma_j^{-1}(x_i - \mu_j)\right)}{\sqrt{(2\pi)^m |\Sigma_j|}},$$

и Σ_j са диагонални матрици.

Твърдение 1. Решението на $\frac{\partial L(\pi,\mu,\Sigma)}{\partial \mu_j} = 0$ има вида $\mu_j = \frac{\sum\limits_{i=1}^{N} \gamma_{ij} x_i}{\sum\limits_{i=1}^{N} \gamma_{ij}}$

Доказателство:

$$0 = \frac{\partial L(\pi,\mu,\Sigma)}{\partial \mu_j} = \sum_{i=1}^{n} \left[\frac{\pi_j \frac{\partial \mathcal{N}(x_i,\mu_j,\Sigma_j)}{\partial \mu_j}}{\sum\limits_{k=1}^{K} \pi_k \mathcal{N}(x_i,\mu_k,\Sigma_k)} \right]$$

Използвайки Свойство 2, можем да намерим производната на $\mathcal{N}(x_i,\mu_j,\Sigma_j)$ по μ_j:

$$\begin{aligned}
\frac{\partial \mathcal{N}(x_i,\mu_j,\Sigma_j)}{\partial \mu_j} &= \partial \left[\frac{exp\left(-\frac{1}{2}(x_i-\mu_j)^T \Sigma_j^{-1}(x_i-\mu_j)\right)}{\sqrt{(2\pi)^m |\Sigma_j|}} \right] / \partial \mu_j \\
&= \frac{exp\left(-\frac{1}{2}(x_i-\mu_j)^T \Sigma_j^{-1}(x_i-\mu_j)\right)}{\sqrt{(2\pi)^m |\Sigma_j|}} \left(-\frac{1}{2} 2\Sigma_j^{-1}(x_i-\mu_j)(-1) \right) \\
&= \mathcal{N}(x_i,\mu_j,\Sigma_j)\Sigma_j^{-1}(x_i-\mu_j)
\end{aligned}$$

Следователно:

$$0 = \frac{\partial L(\pi,\mu,\Sigma)}{\partial \mu_j} = \sum_{i=1}^{n} \left[\frac{\pi_j \frac{\partial \mathcal{N}(x_i,\mu_j,\Sigma_j)}{\partial \mu_j}}{\sum\limits_{k=1}^{K} \pi_k \mathcal{N}(x_i,\mu_k,\Sigma_k)} \right] = \sum_{i=1}^{n} \left[\frac{\pi_j \mathcal{N}(x_i,\mu_j,\Sigma_j)\Sigma_j^{-1}(x_i-\mu_j)}{\sum\limits_{k=1}^{K} \pi_k \mathcal{N}(x_i,\mu_k,\Sigma_k)} \right]$$

$\longleftrightarrow$

$$\sum_{i=1}^{N} \left[\frac{\pi_j \mathcal{N}(x_i,\mu_j,\Sigma_j)\Sigma_j^{-1} x_i}{\sum\limits_{k=1}^{K} \pi_k \mathcal{N}(x_i,\mu_k,\Sigma_k)} \right] = \sum_{i=1}^{N} \left[\frac{\pi_j \mathcal{N}(x_i,\mu_j,\Sigma_j)\Sigma_j^{-1} \mu_j}{\sum\limits_{k=1}^{K} \pi_k \mathcal{N}(x_i,\mu_k,\Sigma_k)} \right]$$

$$\cancel{\Sigma_j^{-1}} \sum_{i=1}^{N} \left[\frac{\pi_j \mathcal{N}(x_i,\mu_j,\Sigma_j) x_i}{\sum\limits_{k=1}^{K} \pi_k \mathcal{N}(x_i,\mu_k,\Sigma_k)} \right] = \cancel{\Sigma_j^{-1}} \sum_{i=1}^{N} \left[\frac{\pi_j \mathcal{N}(x_i,\mu_j,\Sigma_j) \mu_j}{\sum\limits_{k=1}^{K} \pi_k \mathcal{N}(x_i,\mu_k,\Sigma_k)} \right]$$

Нека означим $\gamma_{ij} = \frac{\pi_j \mathcal{N}(x_i,\mu_j,\Sigma_j)}{\sum\limits_{k=1}^{K} \pi_k \mathcal{N}(x_i,\mu_k,\Sigma_k)}$. Тогава имаме:

$$\sum_{i=1}^{N} \gamma_{ij} x_i = \mu_j \sum_{i=1}^{N} \gamma_{ij}$$

$$\mu_j = \frac{\sum\limits_{i=1}^{N} \gamma_{ij} x_i}{\sum\limits_{i=1}^{N} \gamma_{ij}}$$

Твърдение 2. Решението на $\frac{\partial L(\pi, \mu, \Sigma)}{\partial \Sigma_j} = 0$ има вида $\Sigma_j = \begin{cases} \frac{\sum\limits_{i=1}^{N} \gamma_{ij}(x_{it} - \mu_{js})^2}{\sum\limits_{i=1}^{N} \gamma_{ij}}, & t == s \\ 0, & \text{иначе} \end{cases}$

Доказателство:

$$0 = \frac{\partial L(\pi, \mu, \Sigma)}{\partial \Sigma_j} = \sum_{i=1}^{n} \left[\frac{\pi_j \frac{\partial \mathcal{N}(x_i, \mu_j, \Sigma_j)}{\partial \Sigma_j}}{\sum\limits_{k=1}^{K} \pi_k \mathcal{N}(x_i, \mu_k, \Sigma_k)} \right]$$

$\Sigma_j = (\sigma_{ij})_{m \times m}$ и $\sigma_{ij} = 0$, ако $i \neq j$. Първо смятаме:

$$\frac{\partial \left[(x_i - \mu_j)^T \Sigma_j^{-1} (x_i - \mu_j) \right]}{\partial \sigma_{ts}} =$$

$$= \frac{\partial \left(\left((x_{i1} - \mu_{j1}), (x_{i2} - \mu_{j2}), \ldots (x_{im} - \mu_{jm}) \right) \begin{pmatrix} \frac{1}{\sigma_{11}} & 0 & \ldots & 0 \\ 0 & \frac{1}{\sigma_{22}} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \frac{1}{\sigma_{mm}} \end{pmatrix} \begin{pmatrix} (x_{i1} - \mu_{j1}) \\ (x_{i2} - \mu_{j2}) \\ \vdots \\ (x_{im} - \mu_{jm}) \end{pmatrix} \right)}{\partial \sigma_{ts}} =$$

$$= \frac{\partial \left(\frac{(x_{i1} - \mu_{j1})^2}{\sigma_{11}} + \frac{(x_{i2} - \mu_{j2})^2}{\sigma_{22}} + \ldots + \frac{(x_{im} - \mu_{jm})^2}{\sigma_{mm}} \right)}{\partial \sigma_{ts}} =$$

$$= \begin{cases} 0, & t \neq s \\ \frac{-(x_{it} - \mu_{jt})^2}{\sigma_{tt}^2}, & t = s \end{cases}$$

Да разгледаме производната по произволен елемент σ_{ts}:

$$\frac{\partial \mathcal{N}(x_i, \mu_j, \Sigma_j)}{\partial \sigma_{ts}} = \frac{\partial \left[\frac{exp\left(-\frac{1}{2}(x_i - \mu_j)^T \Sigma_j^{-1}(x_i - \mu_t)\right)}{\sqrt{(2\pi)^m |\Sigma_j|}} \right]}{\partial \sigma_{ts}}$$

$$= \frac{\frac{\partial \left[exp\left(-\frac{1}{2}(x_i - \mu_j)^T \Sigma_j^{-1}(x_i - \mu_t)\right)\right]}{\partial \sigma_{ts}} \sqrt{(2\pi)^m |\Sigma_j|} - exp\left(-\frac{1}{2}(x_i - \mu_j)^T \Sigma_j^{-1}(x_i - \mu_t)\right) \frac{\partial \left[\sqrt{(2\pi)^m |\Sigma_j|}\right]}{\partial \sigma_{ts}}}{\left(\sqrt{(2\pi)^m |\Sigma_j|}\right)^2}$$

$$= \frac{exp\left(-\frac{1}{2}(x_i - \mu_j)^T \Sigma_j^{-1}(x_i - \mu_t)\right) \frac{1}{2} \frac{(x_{it} - \mu_{jt})^2}{\sigma_{tt}^2} \sqrt{(2\pi)^m |\Sigma_j|}}{\left(\sqrt{(2\pi)^m |\Sigma_j|}\right)^2}$$

$$- \frac{exp\left(-\frac{1}{2}(x_i - \mu_j)^T \Sigma_j^{-1}(x_i - \mu_t)\right) \frac{1}{2} \frac{(2\pi)^m |\Sigma_j|}{\sigma_{ts} \sqrt{(2\pi)^m |\Sigma_j|}}}{\left(\sqrt{(2\pi)^m |\Sigma_j|}\right)^2} =$$

$$\frac{\mathcal{N}(x_i, \mu_j, \Sigma_j)(x_{it} - \mu_{jt})^2}{2\sigma_{tt}^2} - \frac{\mathcal{N}(x_i, \mu_j, \Sigma_j)}{2\sigma_{tt}} = \frac{\mathcal{N}(x_i, \mu_j, \Sigma_j)}{2\sigma_{tt}^2} \left[(x_{it} - \mu_{jt})^2 - \sigma_{tt}\right]$$

$$0 = \frac{\partial L(\pi, \mu, \Sigma)}{\partial \sigma_{tt}^j} = \sum_{i=1}^{n} \left[\frac{\pi_j \dfrac{\partial \mathcal{N}(x_i, \mu_j, \Sigma_j)}{\partial \sigma_{tt}^j}}{\sum\limits_{k=1}^{K} \pi_k \mathcal{N}(x_i, \mu_k, \Sigma_k)} \right]$$

$$= \sum_{i=1}^{N} \frac{\pi_j \mathcal{N}(x_i, \mu_j, \Sigma_j) \left[(x_{it} - \mu_{jt})^2 - \sigma_{tt} \right]}{2\sigma_{tt}^2 \sum\limits_{k=1}^{K} \pi_k \mathcal{N}(x_i, \mu_k, \Sigma_k)}$$

$\leftrightarrow$

$$\sum_{i=1}^{N} \frac{\pi_j \mathcal{N}(x_i, \mu_j, \Sigma_j)(x_{it} - \mu_{jt})^2}{\cancel{2\sigma_{tt}^2} \sum\limits_{k=1}^{K} \pi_k \mathcal{N}(x_i, \mu_k, \Sigma_k)} = \sum_{i=1}^{N} \frac{\pi_j \mathcal{N}(x_i, \mu_j, \Sigma_j)\sigma_{tt}}{\cancel{2\sigma_{tt}^2} \sum\limits_{k=1}^{K} \pi_k \mathcal{N}(x_i, \mu_k, \Sigma_k)}$$

$$\sum_{i=1}^{N} \gamma_{ij}(x_{it} - \mu_{jt})^2 = \sigma_{tt} \sum_{i=1}^{N} \gamma_{ij}$$

$$\sigma_{tt} = \frac{\sum\limits_{i=1}^{N} \gamma_{ij}(x_{it} - \mu_{jt})^2}{\sum\limits_{i=1}^{N} \gamma_{ij}}$$

$$\Sigma_j = \begin{pmatrix} \dfrac{\sum\limits_{i=1}^{N} \gamma_{ij}(x_{i1} - \mu_{j1})^2}{\sum\limits_{i=1}^{N} \gamma_{ij}} & 0 & \dots & 0 \\ 0 & \dfrac{\sum\limits_{i=1}^{N} \gamma_{ij}(x_{i2} - \mu_{j2})^2}{\sum\limits_{i=1}^{N} \gamma_{ij}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \dfrac{\sum\limits_{i=1}^{N} \gamma_{ij}(x_{im} - \mu_{jm})^2}{\sum\limits_{i=1}^{N} \gamma_{ij}} \end{pmatrix}$$

$$= \begin{cases} \dfrac{\sum\limits_{i=1}^{N} \gamma_{ij}(x_{it} - \mu_{js})^2}{\sum\limits_{i=1}^{N} \gamma_{ij}}, & t == s \\ 0, & \text{иначе} \end{cases}$$

Твърдение 3. Решението на $\frac{\partial L(\pi, \mu, \Sigma)}{\partial \pi_j} = 0$ има вида $\pi_j = \frac{\sum_{i=1}^{N} \gamma_{ij}}{N}$

Доказателство:

$$0 = \frac{\partial L(\pi, \mu, \Sigma)}{\partial \pi_j} = \frac{\partial \left[\sum_{i=1}^{n} log(\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k))) + \lambda(\sum_{k=1}^{K} \pi_k - 1) \right]}{\partial \pi_j} =$$

$$= \sum_{i=1}^{N} \left[\frac{\mathcal{N}(x_i, \mu_j, \Sigma_j)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i, \mu_k, \Sigma_k)} \right] + \lambda = \pi_j \left(\sum_{i=1}^{N} \left[\frac{\mathcal{N}(x_i, \mu_j, \Sigma_j)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i, \mu_k, \Sigma_k)} \right] + \lambda \right)$$

$\longleftrightarrow$

$$-\lambda \pi_j = \pi_j \sum_{i=1}^{N} \left[\frac{\mathcal{N}(x_i, \mu_j, \Sigma_j)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i, \mu_k, \Sigma_k)} \right]$$

$$-\lambda \sum_{j=1}^{K} \pi_j = \sum_{j=1}^{K} \pi_j \sum_{i=1}^{N} \left[\frac{\mathcal{N}(x_i, \mu_j, \Sigma_j)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i, \mu_k, \Sigma_k)} \right]$$

Използваме, че $\sum_{j=1}^{K} \pi_j = 1$ и разменяме местата на сумите в дясно:

$$-\lambda = \sum_{i=1}^{N} \left[\frac{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i, \mu_j, \Sigma_j)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i, \mu_k, \Sigma_k)} \right] = \sum_{i=1}^{N} 1$$

$\lambda = -N$

Заместваме $\lambda = -N$ в

$$-\lambda \pi_j = \pi_j \sum_{i=1}^{N} \left[\frac{\mathcal{N}(x_i, \mu_j, \Sigma_j)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i, \mu_k, \Sigma_k)} \right]$$

$$N\pi_j = \sum_{i=1}^{N} \left[\frac{\pi_j \mathcal{N}(x_i, \mu_j, \Sigma_j)}{\sum_{k=1}^{K} \pi_k \mathcal{N}(x_i, \mu_k, \Sigma_k)} \right]$$

$$N\pi_j = \sum_{i=1}^{N} \gamma_{ij} \longleftrightarrow \pi_j = \frac{\sum_{i=1}^{N} \gamma_{ij}}{N}$$

Приложение Д

Приложение за Максимизиране на ентропията

Нека имаме входни данни $\mathcal{D} = (x_1, y_1), ..., (x_n, y_n)$, където $x_i \in X$, а $y_i \in Y$, X и Y - изброими.

Търсим това разпределение p, което приближава разпределението, генерирало данните в $\mathcal{D}$, и не прави допълнителни предположения извън $\mathcal{D}$.

Тоест търсеното разпределение p, трябва да изпълнява:

$p(x, y) = \tilde{p}(x)p(y|x),$

Тук с $\tilde{p}$ означаваме емпиричното разпределение, дефинирано като:

$\forall (x, y) \in X \times Y : \tilde{p}(x, y) = \frac{\#(x, y)}{n}$, където $\#(x, y) :=$ брой срещания на (x, y) в $\mathcal{D}$.

С други думи:

$$\begin{aligned} p(x) = \sum_{y \in Y} p(x, y) &= \sum_{y \in Y} \tilde{p}(x)p(y|x) \\ &= \tilde{p}(x) \sum_{y \in Y} p(y|x) \\ &= \tilde{p}(x) \end{aligned}$$

и искаме да максимизира ентропията:

$$H_p(X, Y) = - \sum_{(x,y) \in X \times Y} p(x, y)\, log(p(x, y))$$

Тъй като:

$\sum_{x \in Proj_1(\mathcal{D})} p(x) = \sum_{x \in Proj_1(\mathcal{D})} \tilde{p}(x) = 1 = \sum_{x \in Proj_1(\mathcal{D})} \tilde{p}(x) = \sum_{x \in X} \tilde{p}(x) = \sum_{x \in X} p(x),$ то следва, че $\sum_{x \notin Proj_1(\mathcal{D})} p(x) = 0,$

тогава е достатъчно да искаме $p(x) = \tilde{p}(x)$ само за $x \in Proj_1(\mathcal{D})$

Нека имаме още множество от характеристични функции $\mathcal{H}, |\mathcal{H}| = K$, които са от вида $h_i : X \times Y \to [0, 1]$.

Ако с $E(q, h)$ означим очакването на h, спрямо разпределение q, тоест:

$$E(q, h) = \sum_{(x,y)\in X\times Y} q(x, y)h(x, y)$$

То искаме за търсеното p да е изпълнено:

$$E(p, h) = E(\tilde{p}, h), \forall h \in \mathcal{H}$$

Ако дефинираме допълнително функции $h_{x_0}(x, y) = \begin{cases} 1 & x = x_0 \\ 0 & \text{иначе} \end{cases}$, то можем да изразим $p(x) = \tilde{p}(x)$ по следния начин:

$$p(x_0) = \sum_{y\in Y} p(x_0, y) = \sum_{(x,y)\in X\times Y} p(x, y)h_{x_0}(x, y) = E(p, h_{x_0})$$
$$\forall x \in Proj_1(\mathcal{D}) : E(p, h_x) = E(\tilde{p}, h_x)$$

Дефиниция. (Множество от допустими вероятностни разпределения)

$$P = \{p| \, (\forall x \in Proj_1(\mathcal{D}) : E(p, h_x) = E(\tilde{p}, h_x)) \wedge (\forall h \in \mathcal{H} : E(p, h) = E(\tilde{p}, h))\} \quad ,$$

тогава искаме да намерим

$$\begin{aligned}
\hat{p} &= argmax_{p\in P} H_p(X, Y) \\
&= argmax_{p\in P} \left(- \sum_{(x,y)\in X\times Y} p(x, y) \, log(p(x, y)) \right) \\
&= argmax_{p\in P} \left(- \sum_{(x,y)\in X\times Y} p(x, y) \, log(\tilde{p}(x)p(y|x)) \right) \\
&= argmax_{p\in P} \left(- \sum_{(x,y)\in X\times Y} p(x, y) log(\tilde{p}(x)) - \sum_{(x,y)\in X\times Y} p(x, y) \, log(p(y|x)) \right) \\
&= argmax_{p\in P} \left(- \sum_{(x,y)\in X\times Y} \tilde{p}(x)p(y|x) log(\tilde{p}(x)) + H_p(Y|X) \right) \\
&= argmax_{p\in P} \left(- \sum_{x\in X} \tilde{p}(x) log(\tilde{p}(x)) \sum_{y\in Y} p(y|x) + H_p(Y|X) \right) \\
&= argmax_{p\in P} \left(- \sum_{x\in X} \tilde{p}(x) log(\tilde{p}(x)) + H_p(Y|X) \right)
\end{aligned}$$

$- \sum_{x\in X} \tilde{p}(x) log(\tilde{p}(x))$ е константа спрямо p, следователно:

$$= argmax_{p\in P} H_p(Y|X)$$

За да решим тази оптимизационна задача, ще ползваме множители на Лагранж. Тъй като имаме K ограничения за всяка от характеристичните функции и трябва да отчетем, че търсим разпределение с определени свойства,

задачата ще има вида:

$$\Lambda(p,\tau,\lambda,\mu) = H_p(X,Y) + \sum_{x\in Proj_1(\mathcal{D})} \tau_x \left(E(p,h_x) - E(\tilde{p},h_x)\right) + \sum_{i=1}^{K} \lambda_i (E(p,h_i) - E(\tilde{p},h_i)) + \mu \left[\sum_{(x,y)\in X\times Y} p(x,y) - 1\right]$$

Нека фиксираме едно $x_0 \in X, y_0 \in Y$.

$$\frac{\partial\left(\Lambda(p,\tau,\lambda,\mu)\right)}{\partial p(x_0,y_0)} = \frac{\partial H_p(X,Y)}{\partial p(x_0,y_0)} + \frac{\partial\left(\sum_{x\in Proj_1(\mathcal{D})} \tau_x \left(E(p,h_x) - E(\tilde{p},h_x)\right)\right)}{\partial p(x_0,y_0)} + \frac{\partial\left(\sum_{i=1}^{K} \lambda_i (E(p,h_i) - E(\tilde{p},h_i))\right)}{\partial p(x_0,y_0)} + \frac{\partial\left(\mu\left[\sum_{(x,y)\in X\times Y} p(x,y) - 1\right]\right)}{\partial p(x_0,y_0)}$$

$$= \frac{\partial\left(-\sum_{(x,y)\in X\times Y} p(x,y)\, log(p(x,y))\right)}{\partial p(x_0,y_0)} + \frac{\partial\left(\sum_{x\in Proj_1(\mathcal{D})} \tau_x \left[\sum_{(x',y)\in X\times Y} p(x',y) h_x(x',y)\right]\right)}{\partial p(x_0,y_0)} + \frac{\partial\left(\sum_{i=1}^{K} \lambda_i \left[\sum_{(x,y)\in X\times Y} p(x,y) h_i(x,y)\right]\right)}{\partial p(x_0,y_0)} + \mu$$

$$= -log(p(x_0,y_0)) - 1 + \sum_{x\in Proj_1(\mathcal{D})} \tau_x h_x(x_0,y_0) + \sum_{i=1}^{K} \lambda_i h_i(x_0,y_0) + \mu$$

$$= -log(p(x_0,y_0)) - 1 + \sum_{i=1}^{K'} \lambda_i h_i(x_0,y_0) + \mu,$$

където $K' = K + |Proj_1(\mathcal{D})|$ и ако номерираме x-овете в $Proj_1(\mathcal{D})$ от $K+1$ до K', то $\forall j = K+1,\dots,K' : \lambda_j = \tau_{x_j}$, а $h_j = h_{x_j}$

Искаме да нулираме производната:

$$-log(p(x_0,y_0)) - 1 + \sum_{i=1}^{K'} \lambda_i h_i(x_0,y_0) + \mu = 0 \longleftrightarrow$$

$$log(p(x_0,y_0)) = \sum_{i=1}^{K'} \lambda_i h_i(x_0,y_0) + \mu - 1 \longleftrightarrow$$

$$p(x_0,y_0) = exp\left(\sum_{i=1}^{K'} \lambda_i h_i(x_0,y_0)\right) exp\,(\mu - 1) \qquad (Д.0.1)$$

Производната по μ ни дава:

$$\sum_{(x,y)\in X\times Y} p(x,y) = 1 \longleftrightarrow$$

$$\sum_{(x,y)\in X\times Y} exp\left(\sum_{i=1}^{K'} \lambda_i h_i(x,y) + \mu - 1\right) = 1 \longleftrightarrow$$

$$exp\,(\mu-1) \sum_{(x,y)\in X\times Y} exp\left(\sum_{i=1}^{K'} \lambda_i h_i(x,y)\right) = 1$$

$$\longleftrightarrow$$

$$exp\,(\mu-1) = \frac{1}{\sum\limits_{(x,y)\in X\times Y} exp\left(\sum\limits_{i=1}^{K'} \lambda_i h_i(x,y)\right)}$$

Заместваме в Уравнение Д.0.1:

$$p(x_0,y_0) = \frac{exp\left(\sum\limits_{i=1}^{K'} \lambda_i h_i(x_0,y_0)\right)}{\sum\limits_{(x,y)\in X\times Y} exp\left(\sum\limits_{i=1}^{K'} \lambda_i h_i(x,y)\right)}$$

Тъй като $p(x,y) = p(x)p(y|x)$, можем да получим, че:

$$p(y_0|x_0) = \frac{p(x_0,y_0)}{p(x_0)} = \frac{p(x_0,y_0)}{\sum\limits_{y\in Y} p(x_0,y)}$$

$$= \frac{exp\left(\sum\limits_{i=1}^{K'} \lambda_i h_i(x_0,y_0)\right)}{\sum\limits_{(x',y')\in X\times Y} exp\left(\sum\limits_{i=1}^{K'} \lambda_i h_i(x',y')\right)} \div \sum_{y\in Y}\left(\frac{exp\left(\sum\limits_{i=1}^{K'} \lambda_i h_i(x_0,y)\right)}{\sum\limits_{(x',y')\in X\times Y} exp\left(\sum\limits_{i=1}^{K'} \lambda_i h_i(x',y')\right)}\right)$$

$$= \frac{exp\left(\sum\limits_{i=1}^{K'} \lambda_i h_i(x_0,y_0)\right)}{\cancel{\sum\limits_{(x',y')\in X\times Y} exp\left(\sum\limits_{i=1}^{K'} \lambda_i h_i(x',y')\right)}} \div \frac{\sum\limits_{y\in Y} exp\left(\sum\limits_{i=1}^{K'} \lambda_i h_i(x_0,y)\right)}{\cancel{\sum\limits_{(x',y')\in X\times Y} exp\left(\sum\limits_{i=1}^{K'} \lambda_i h_i(x',y')\right)}}$$

$$= \frac{exp\left(\sum\limits_{i=1}^{K'} \lambda_i h_i(x_0,y_0)\right)}{\sum\limits_{y\in Y} exp\left(\sum\limits_{i=1}^{K'} \lambda_i h_i(x_0,y)\right)} = \frac{exp\left(\sum\limits_{i=1}^{K} \lambda_i h_i(x_0,y_0)\right) \cancel{exp\left(\sum\limits_{i=K+1}^{K'} \lambda_i h_i(x_0,y_0)\right)}}{\cancel{exp\left(\sum\limits_{i=K+1}^{K'} \lambda_i h_i(x_0,y)\right)} \sum\limits_{y\in Y} exp\left(\sum\limits_{i=1}^{K} \lambda_i h_i(x_0,y)\right)}$$

$$= \frac{exp\left(\sum\limits_{i=1}^{K} \lambda_i h_i(x_0,y_0)\right)}{\sum\limits_{y\in Y} exp\left(\sum\limits_{i=1}^{K} \lambda_i h_i(x_0,y)\right)},$$

тъй като за $i = K+1, \dots, K', h_i(x,y)$ не зависят от избора на y.

Следователно вида на търсеното $\hat{p}$ е $\hat{p}(x,y) = \pi \prod_{i=1}^{K'} e^{\lambda_i h_i(x,y)}$, като π е нормализиращата константа. Ще покажем че $\hat{p}$, което максимизира ентропията, също максимизира и условното правдоподобие.

Нека с Q означим всички разпределения с желания вид.
Дефиниция. (Множество от вероятностни разпределения с търсения вид)

$$Q = \{p \mid p(x,y) = \pi \prod_{i=1}^{K'} e^{\lambda_i h_i(x,y)}\} \quad ,$$

За да намерим оптималното разпределение, ще ни е нужно да дефинираме разстояние между разпределения - "Разстояние" на Кулбек-Лайблър:

$$D(p,q) = \sum_{(x,y)\in X\times Y} p(x,y)\, log\left(\frac{p(x,y)}{q(x,y)}\right)$$

"Разстоянието" на Кулбек-Лайблър всъщност не е разстояние в математическия смисъл (в смисъла на метрика), тъй като не е симетрична функция, но често се използва за сравнение на разпределения, тъй като има този интуитивен смисъл. Затова ще продължим да го наричаме разстояние, пропускайки кавичките.

С това сме готови да покажем следните твърдения:
Твърдение 4. За всеки две разпределения p и q върху $X \times Y$, $D(p,q) \geq 0$, като $D(p,q) = 0 \iff p = q$

Доказателство: Тъй като p е разпределение и е изпълнено, че $\sum_{(x,y)\in X\times Y} p(x,y) = 1$, можем да приложим неравенството на Йенсен:

$$\sum_{i=1}^{\infty} p(x_i,y_i) f(z_i) \leq f\left(\sum_{i=1}^{\infty} p(x_i,y_i) z_i\right), \forall i : z_i \in \mathbb{R},$$

където f е вдлъбната. Ако за f е изпълнено, че $f'' < 0$, то равенство се достига, когато $\forall i,j : z_i = z_j$.

$$\begin{aligned}
-D(p,q) &= -\sum_{(x,y)\in X\times Y} p(x,y)\, log\left(\frac{p(x,y)}{q(x,y)}\right) \\
&= \sum_{(x,y)\in X\times Y} p(x,y)\, log\left(\frac{q(x,y)}{p(x,y)}\right) \\
&\overset{\text{Йенсен}}{\leq} log\left(\sum_{(x,y)\in X\times Y} \cancel{p(x,y)} \frac{q(x,y)}{\cancel{p(x,y)}}\right) \\
&= log\left(\sum_{(x,y)\in X\times Y} q(x,y)\right) = 0 \\
&\longleftrightarrow D(p,q) \geq 0
\end{aligned}$$

Тъй като втората производна на логаритъма е винаги отрицателна, равенство при неравенството на Йенсен се достига, когато $\frac{q(x,y)}{p(x,y)}$ е константа, тоест:

$$q(x,y) = Cp(x,y)$$
$$\sum_{(x,y)\in\mathcal{D}} q(x,y) = \sum_{(x,y)\in\mathcal{D}} Cp(x,y)$$
$$\longleftrightarrow C = 1$$
$$\longleftrightarrow p(x,y) = q(x,y) \forall (x,y) \in X \times Y$$

Твърдение 5. За всеки $p_1, p_2 \in P, q \in Q$ е изпълнено:

$$\sum_{(x,y)\in X\times Y} p_1(x,y)\, log(q(x,y)) = \sum_{(x,y)\in X\times Y} p_2(x,y)\, log(q(x,y))$$

Доказателство:

$$\sum_{(x,y)\in X\times Y} p_1(x,y)\, log(q(x,y))$$

$$= \sum_{(x,y)\in X\times Y} p_1(x,y)\, log\left(\pi \prod_{i=1}^{K'} e^{\lambda_i h_i(x,y)}\right)$$

$$= \sum_{(x,y)\in X\times Y} p_1(x,y)\left(log(\pi) + log\left(\prod_{i=1}^{K'} e^{\lambda_i h_i(x,y)}\right)\right)$$

$$= \sum_{(x,y)\in X\times Y} p_1(x,y)\left(log(\pi) + \sum_{i=1}^{K'} log\left(e^{\lambda_i h_i(x,y)}\right)\right)$$

$$= \sum_{(x,y)\in X\times Y} p_1(x,y)\left(log(\pi) + \sum_{i=1}^{K'} \lambda_i h_i(x,y)\right)$$

$$= \sum_{(x,y)\in X\times Y} p_1(x,y)\, log(\pi) + \sum_{(x,y)\in X\times Y} p_1(x,y) \sum_{i=1}^{K'} \lambda_i h_i(x,y)$$

$$= log(\pi) \sum_{(x,y)\in X\times Y} p_1(x,y) + \sum_{(x,y)\in X\times Y} \sum_{i=1}^{K'} p_1(x,y)\lambda_i h_i(x,y)$$

$$= log(\pi).1 + \sum_{(x,y)\in X\times Y} \sum_{i=1}^{K'} p_1(x,y)\lambda_i h_i(x,y)$$

$$= log(\pi).1 + \sum_{i=1}^{K'} \lambda_i \sum_{(x,y)\in X\times Y} p_1(x,y) h_i(x,y)$$

$$= log(\pi).1 + \sum_{i=1}^{K'} \lambda_i E(p_1, h_i)$$

Тъй като $p_2 \in P$ и $E(p_1,h) = E(\tilde{p},h_i) = E(p_2,h_i) \forall i = 1,\dots,K'$:

$$= log(\pi).1 + \sum_{i=1}^{K'} \lambda_i E(p_2, h_i)$$

$$= log(\pi).1 + \sum_{i=1}^{K'} \lambda_i \sum_{(x,y)\in X\times Y} p_2(x,y) h_i(x,y)$$

Използваме и че $\sum_{(x,y)\in X\times Y} p_2(x,y) = 1$

$$= log(\pi) \sum_{(x,y)\in X\times Y} p_2(x,y) + \sum_{i=1}^{K'} \lambda_i \sum_{(x,y)\in X\times Y} p_2(x,y) h_i(x,y)$$

$$= \sum_{(x,y)\in X\times Y} p_2(x,y)\, log(q(x,y))$$

Твърдение 6. Ако $p \in P, q \in Q, r \in P \cap Q$, то $D(p,q) = D(p,r) + D(r,q)$

Доказателство:

$$D(p,r) + D(r,q) =$$

$$= \sum_{(x,y)\in X\times Y} p(x,y)\, log\left(\frac{p(x,y)}{r(x,y)}\right) + \sum_{(x,y)\in X\times Y} r(x,y)\, log\left(\frac{r(x,y)}{q(x,y)}\right)$$

$$= \sum_{(x,y)\in X\times Y} p(x,y)\, log(p(x,y)) - \sum_{(x,y)\in X\times Y} p(x,y)\, log(r(x,y)) +$$

$$\sum_{(x,y)\in X\times Y} r(x,y)\, log(r(x,y)) - \sum_{(x,y)\in X\times Y} r(x,y)\, log(q(x,y))$$

по Твърдение 5 за $p, r \in P$ и $r \in Q$

$$= \sum_{(x,y)\in X\times Y} p(x,y)\, log(p(x,y)) - \cancel{\sum_{(x,y)\in X\times Y} p(x,y)\, log(r(x,y))} +$$

$$\cancel{\sum_{(x,y)\in X\times Y} p(x,y)\, log(r(x,y))} - \sum_{(x,y)\in X\times Y} r(x,y)\, log(q(x,y))$$

$$= \sum_{(x,y)\in X\times Y} p(x,y)\, log(p(x,y)) - \sum_{(x,y)\in X\times Y} r(x,y)\, log(q(x,y))$$

по Твърдение 5 за $p, r \in P$ и $q \in Q$

$$= \sum_{(x,y)\in X\times Y} p(x,y)\, log(p(x,y)) - \sum_{(x,y)\in X\times Y} p(x,y)\, log(q(x,y))$$

$$= \sum_{(x,y)\in X\times Y} p(x,y)\, log\left(\frac{p(x,y)}{q(x,y)}\right) = D(p,q)$$

Твърдение 7. Ако $r \in P \cap Q$, то r е единствено и $r = argmax_{p\in P} H_p(X,Y)$

Доказателство:

Нека $r \in P \cap Q$. Ще покажем, че за всяко $p \in P$: $H_r(X,Y) \geq H_p(X,Y)$.

Нека $u \in Q$, такова че $u(x,y) \neq 0, \forall (x,y) \in X \times Y$. Съществост всяко разпределение q от Q е такова, защото $\sum_{(x,y)\in X\times Y} e^{\bullet} > 0$, а $\pi \neq 0$, защото π е константа и ако $\pi = 0$, тогава $\sum_{(x,y)\in X\times Y} q(x,y) = 0$ и не изпълнява условието за разпределение.

Нека фиксираме произволно $p \in P$. Тогава от Твърдение 6 следва, че

$$D(p,u) = D(p,r) + D(r,u)$$

$$D(p,u) \overset{\text{Твърдение 4}}{\geq} D(r,u)$$

$$\sum_{(x,y)\in X\times Y} p(x,y)\, log\left(\frac{p(x,y)}{u(x,y)}\right) \geq \sum_{(x,y)\in X\times Y} r(x,y)\, log\left(\frac{r(x,y)}{u(x,y)}\right)$$

$$- H_p(X,Y) - \sum_{(x,y)\in X\times Y} p(x,y)\, log(u(x,y)) \geq -H_r(X,Y) - \sum_{(x,y)\in X\times Y} r(x,y)\, log(u(x,y))$$

по Твърдение 5 за $p, r \in P$ и $u \in Q$ следва :

$$- H_p(X,Y) - \cancel{\sum_{(x,y)\in X\times Y} p(x,y)\, log(u(x,y))} \geq -H_r(X,Y) - \cancel{\sum_{(x,y)\in X\times Y} p(x,y)\, log(u(x,y))}$$

$$H_r(X,Y) \geq H_p(X,Y)$$

Следователно $r = argmax_{p \in P} H_p(X, Y)$

Сега нека видим защо r е единствено. Нека $r' = argmax_{p \in P} H_p(X, Y)$. Тогава:

$H_{r'}(X, Y) = H_r(X, Y) \longleftrightarrow D(r, u) = D(r', u)$
но $D(r, u) = D(r, r') + D(r', u)$ по Твърдение 6
$\implies \quad D(r, r') = 0$
$\overset{\text{Твърдение 4}}{\implies} \quad r = r'$

Дефинираме функцията $L(p)$:

$$L(p) = \sum_{(x,y) \in X \times Y} \tilde{p}(x, y)\, log(p(x, y))$$

Твърдение 8. Ако $r \in P \cap Q$, то $r = argmax_{q \in Q} L(q)$

Доказателство: Искаме да покажем, че за всяко $q \in Q : L(r) \geq L(q)$.

Нека фиксираме едно $q \in Q$, а $\tilde{p}$ е емпиричното разпределение и следователно $\tilde{p} \in P$ по дефиниция.

Тогава от Твърдение 6 следва, че:

$D(\tilde{p}, q) = D(\tilde{p}, r) + D(r, q)$
$D(\tilde{p}, q) \overset{\text{Твърдение 4}}{\geq} D(\tilde{p}, r)$
$$\sum_{(x,y) \in X \times Y} \tilde{p}(x, y) log \left(\frac{\tilde{p}(x, y)}{q(x, y)} \right) \geq \sum_{(x,y) \in X \times Y} \tilde{p}(x, y) log \left(\frac{\tilde{p}(x, y)}{r(x, y)} \right)$$
$- \cancel{H_{\tilde{p}}(X, Y)} - L(q) \geq - \cancel{H_{\tilde{p}}(X, Y)} - L(r)$
$\longleftrightarrow L(r) \geq L(q)$

Дефиницията на условно правдоподобие на разпределение p при дадено множество $\mathcal{D}$ е следната:

$$\widehat{L}_{\mathcal{D}}(Y|X) = \prod_{(x,y) \in X \times Y} p(y|x)^{\#(x,y)}$$

Тъй като логаритъмът е вдлъбната и монотонно растяща функция, често се разглежда за удобство:

$$log\left(\widehat{L}_{\mathcal{D}}(Y|X)\right) = \sum_{(x,y) \in X \times Y} \#(x, y) log(p(y|x))$$

Тъй като $\hat{p} \in P \cap Q$, по горното твърдение имаме:

$$
\begin{aligned}
\hat{p} &= argmax_{p \in Q} L(p) \\
&= argmax_{p \in Q} \left(\sum_{(x,y) \in X \times Y} \tilde{p}(x,y) log(p(x,y)) \right) \\
&= argmax_{p \in Q} \left(\sum_{(x,y) \in X \times Y} \tilde{p}(x,y) log(\tilde{p}(x) p(y|x)) \right) \\
&= argmax_{p \in Q} \left(\sum_{(x,y) \in X \times Y} \tilde{p}(x,y) log(\tilde{p}(x)) + \sum_{(x,y) \in X \times Y} \tilde{p}(x,y) log(p(y|x)) \right) \\
&= argmax_{p \in Q} \left(\sum_{(x,y) \in X \times Y} \tilde{p}(x,y) log(p(y|x)) \right) \\
&= argmax_{p \in Q} \left(0 + \sum_{(x,y) \in \mathcal{D}} \frac{\#(x,y)}{n} log(p(y|x)) \right) \\
&= argmax_{p \in Q} \left(\sum_{(x,y) \in \mathcal{D}} \#(x,y) log(p(y|x)) \right) \\
&= argmax_{p \in Q} \left(log \left(\widehat{L}_{\mathcal{D}}(Y|X) \right) \right)
\end{aligned}
$$

От Твърдение 7 и Твърдение 8, че ако вземем разпределение от сечението на P и Q, то е единствено и е равно на $\hat{p} = argmax_{p \in P} H_p(X,Y) = argmax_{p \in P} H_p(Y|X) = argmax_{q \in Q} L(q) = argmax_{q \in Q} log \left(\widehat{L}_{\mathcal{D}}(Y|X) \right)$

Библиография

[ACC19] Gopala Anumanchipalli, Josh Chartier и Edward Chang. “Speech synthesis from neural decoding of spoken sentences”. В: Nature 568 (апр. 2019), с. 493—498. DOI: 10.1038/s41586-019-1119-1.

[AF17] S. M. Alarcão и M. J. Fonseca. “Emotions Recognition Using EEG Signals: A Survey”. В: IEEE Transactions on Affective Computing 10.3 (2017), с. 374—393. DOI: 10.1109/TAFFC.2017.2714671.

[AV07] David Arthur и Sergei Vassilvitskii. “K-means++: the advantages of careful seeding”. В: In Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms. 2007.

[Bar17] Lisa Feldman Barrett. How Emotions Are Made: The Secret Life of the Brain. 2017.

[Bis06] Christopher M. Bishop. Pattern Recognition and Machine Learning. Springer, 2006.

[BSS00] Gary Bell, Marian Schultz и James Schultz. “Voice Recognition in Fighter Aircraft”. В: Journal of Aviation/Aerospace Education & Research (ян. 2000). DOI: 10.15394/jaaer.2000.1270.

[BTV] BTV. URL: https://web.archive.org/web/20190714122441/https://btvnovinite.bg/predavania/tazi-sutrin (дата на посещ. 14.07.2019).

[Bur+05] Felix Burkhardt и др. “A database of German emotional speech”. В: т. 5. Ян. 2005, с. 1517—1520.

[Car21] Jung Carl. Psychological Types. Rascher Verlag, 1921.

[EKK] Moataz El Ayadi, Mohamed S. Kamel и Fakhri Karray. “Survey on Speech Emotion Recognition: Features, Classification Schemes, and Databases”. В: Pattern Recogn. (). DOI: 10.1016/j.patcog.2010.09.020.

[Gha+17] M. Ghai и др. “Emotion recognition on speech signals using machine learning”. В: 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC). 2017. DOI: 10 . 1109 / ICBDACI . 2017 . 8070805.

[KNS09] Shashidhar Koolagudi, Sourav Nandy и K Sreenivasa Rao. “Spectral Features for Emotion Classification”. В: апр. 2009, с. 1292—1296. DOI: 10.1109/IADCC.2009.4809202.

[Meh74] J. A. Mehrabian A. & Russell. An approach to environmental psychology. MIT Press, 1974, с. 216—217.

[OWN96] Alan V. Oppenheim, Alan S. Willsky и S. Hamid Nawab. Signals &Amp; Systems (2Nd Ed.) Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1996.

[Plu01] Robert Plutchik. “The Nature of Emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice”. В: American Scientist 89 (юли 2001), с. 344—350.

[Qua01] Thomas Quatieri -. Discrete-time Speech Signal Processing: Principles and Practice. First. Upper Saddle River, NJ, USA: Prentice Hall Press, 2001.

[RS78] L. Rabiner и R. Schafer. Digital Processing of Speech Signals. Englewood Cliffs: Prentice Hall, 1978.

[SCC11] P. Shen, Z. Changjun и X. Chen. "Automatic Speech Emotion Recognition using Support Vector Machine". В: Proceedings of 2011 International Conference on Electronic Mechanical Engineering and Information Technology. Т. 2. 2011. DOI: 10.1109/EMEIT.2011.6023178.

[Tay09] Paul Taylor. Text-to-Speech Synthesis. 1st. New York, NY, USA: Cambridge University Press, 2009.

[VA06] Thurid Vogt и Elisabeth André. "Improving Automatic Emotion Recognition from Speech via Gender Differentiation". В: Proc. Language Resources and Evaluation Conference (LREC 2006). Genoa, 2006.

[Wee+00] B. Weedon и др. "Perceived Urgency in Speech Warnings". В: Proceedings of the Human Factors and Ergonomics Society Annual Meeting 44 (юли 2000), с. 690—693. DOI: 10.1177/154193120004402251.

[ZZL16] Wei-Long Zheng, Jia-Yi Zhu и Bao-Liang Lu. "Identifying Stable Patterns over Time for Emotion Recognitionsurvey from EEG". В: IEEE Transactions on Affective Computing (ян. 2016). DOI: 10.1109/TAFFC.2017.2712143.

[Тал66] Димитър Талев. Гласовете ви чувам. 1966.