

Дипломна работа

Диана Генева <dageneva@qtrp.org>

2019

Съдържание

1	Нулева зона	3
2	Емоции в реч	4
2.1	Физика на тъгата	4
2.2	Приближаване с тръби	7
2.2.1	Преминаване от една тръба в друга	9
2.2.2	Ограничения при устните	12
2.2.3	Ограничения при глотиса	13
2.2.4	Общ вид на \mathcal{V}	15
2.2.5	Общ вид на \mathcal{X}	17
2.2.6	Общ вид на \mathcal{G}	17
2.2.7	Общ вид на \mathcal{Y}	18
2.3	Представяне със системи	19
2.4	Характеристики	21
2.4.1	Избор	21
2.4.2	Извличане	23
2.5	Класификация	25
2.6	Резултати	25
3	Грубо в мозъка	26
3.1	Характеристики	26
3.1.1	Избор	26
3.1.2	Извличане	26
3.2	Класификация	26
3.3	Резултати	26
4	Двойната звезда	27
4.1	Резултати	27
5	Големият портрет	28
A	Фурие приложение	29
A.1	Дефиниция	29
A.2	Свойства	31
A.3	Конволюция	32
Б	Приложение за полюси и нули	34
Б.1	Дефиниция	34
Б.2	Характеризация на филтри	35

В Приложение към Емоции в реч	40
Библиография	42

Глава 1

Нулева зона

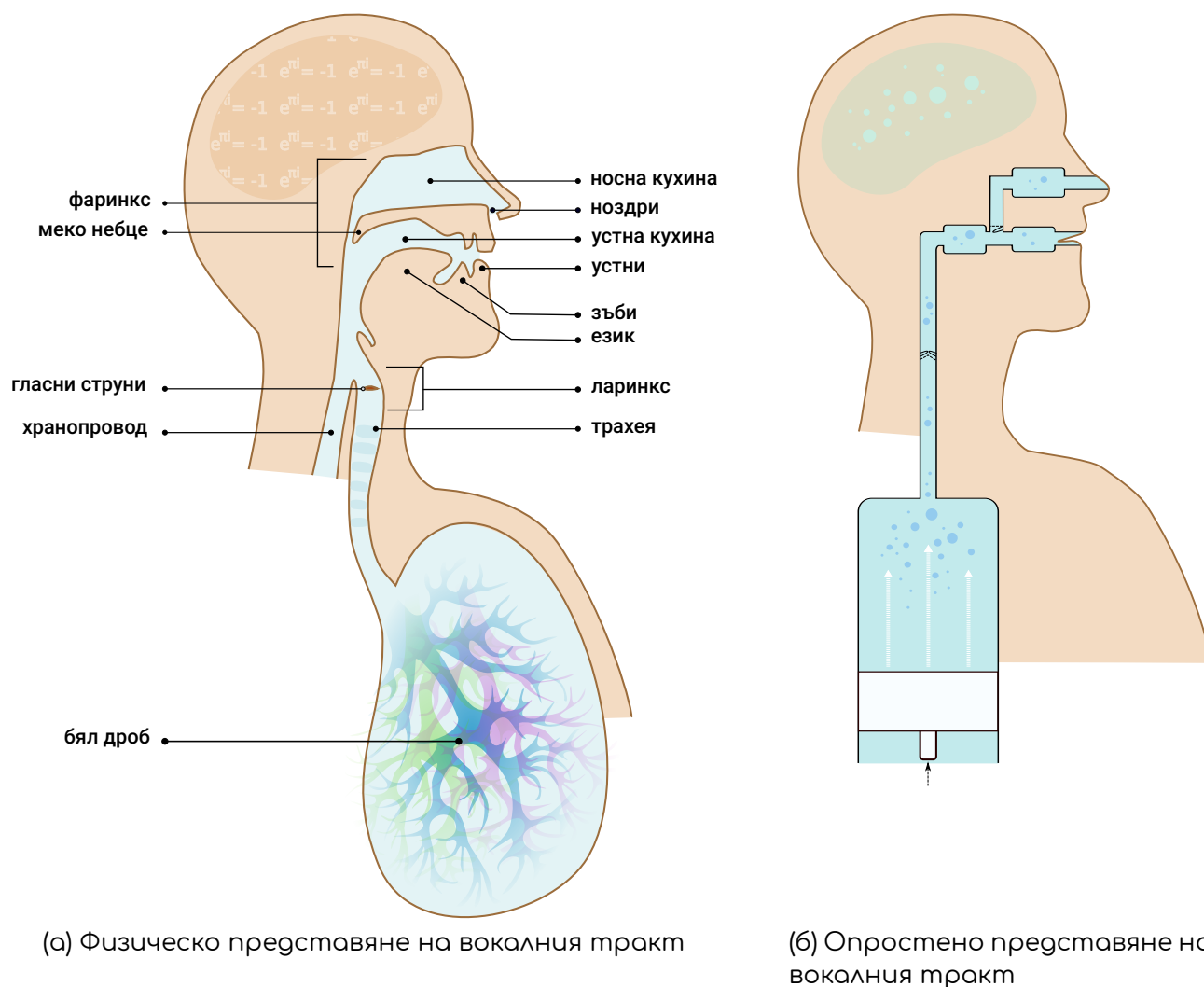
Бла, бла, бла, аз съм толкова емоционална. Не знам



Глава 2

Емоции в реч

2.1 Физика на тъгата



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 2.1.1a. 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 извлечем характеристики, които описват тази конфигурация.

Да разгледаме по-подробно Фигура 2.1.1b, 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 изтласквания от белия дроб въздух.

Белият дроб работи като енергиен източник за тази системата - въздушният поток, получен при свиването му от междуребрните мускули и диафрагмата, се пропъгува 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 пропъгувания 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)

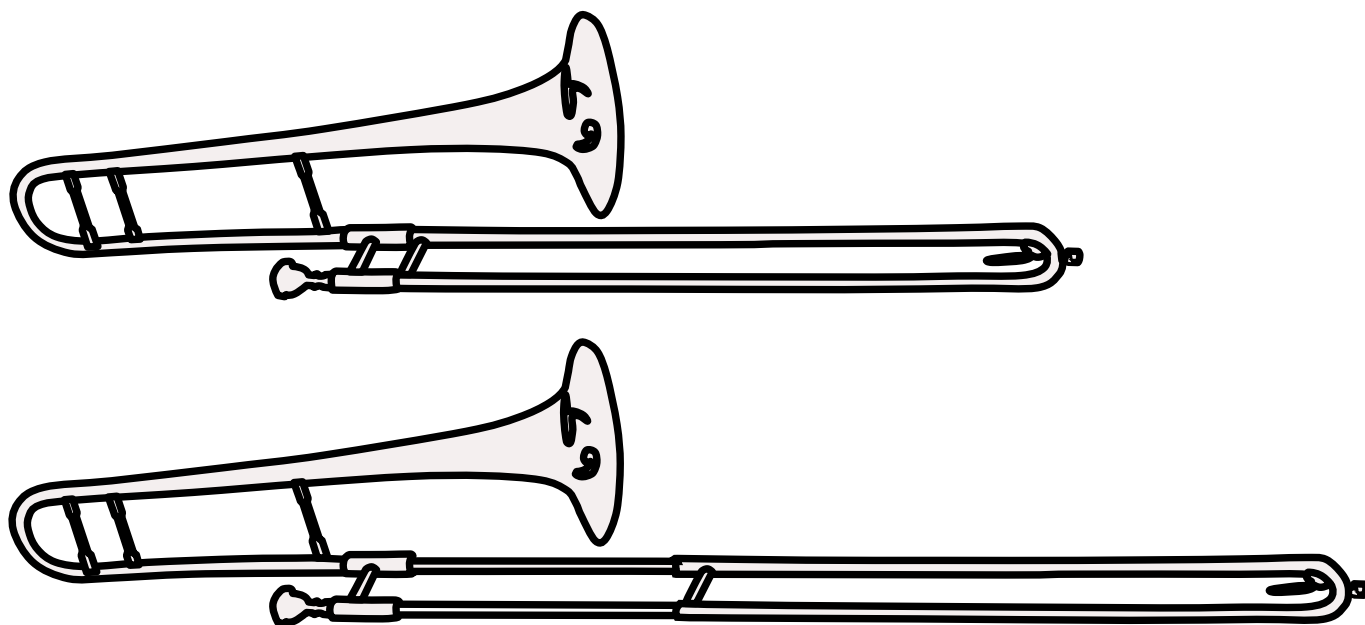
При образуването на проходни звуци, вълната среща преграда по пътя си (като например зъби, устни) и се получава турбуленция, при опита да бъде избутан въздухът през преградата.

3. Преградни (експлозивни)

Те се получават при напълно затворена преградата, за която се натрупа 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 фонemi. 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. Тогав 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 свиренето на духови инструменти.



Фигура 2.1.2: Тромбон

При тях по някакъв начин се променя изходът на вълната, например отпущване и запущване 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 тръбата, 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 пропъгува 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 тракт¹, 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. 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 $g \xleftrightarrow{\mathcal{FS}} G(z)$, $v \xleftrightarrow{\mathcal{FS}} V(z)$, $r \xleftrightarrow{\mathcal{FS}} R(z)$, а новият сигнал е y с $y \xleftrightarrow{\mathcal{FS}} Y(z)$, е изпълнено, че

$$Y(z) = G(z)V(z)R(z),$$

където $z = e^{i\omega_k}$ е прост сигнал с ъглова честота ω_k , а $\xleftrightarrow{\mathcal{FS}}$ обозначава Фурие

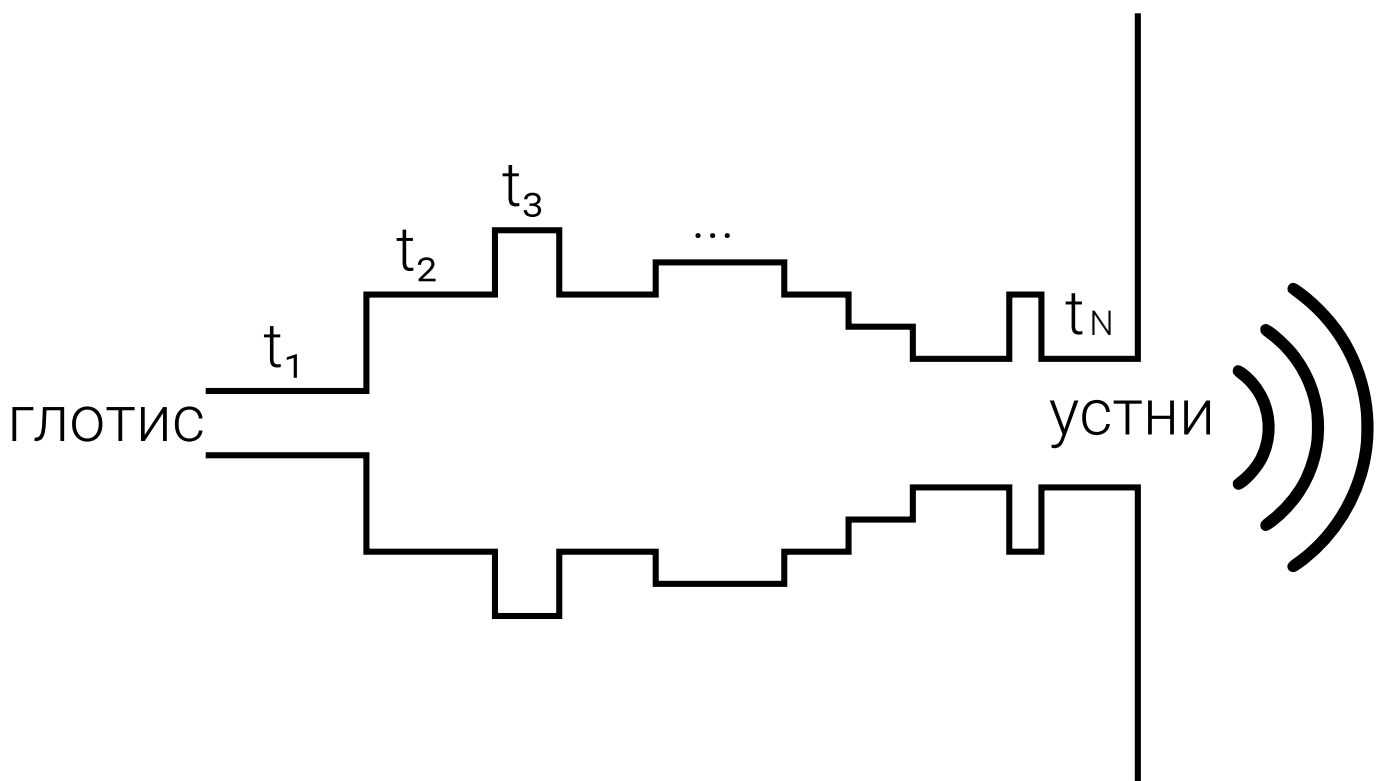
¹както се зарекохме в предния раздел

преобразуване, което е въведено в [Приложение А](#).

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 усните, 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 g , който ще бъде променен от филтъра на вокалния тракт.

Бележка: Дефиницията за Фурие преобразуване изисква сингалът 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

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 $1 \dots N$, 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 промяната в обемната скорост на позиция x в момента t

5. $p = p(x, t)$ - е промяната в звуковото налягане

Звуковите вълни, преминаващи през течна среда в тръба, изпълняват уравнения:

$$\frac{\partial \rho}{\partial x} = \frac{\rho}{A} \frac{\partial u}{\partial t} \quad (2.2.1a)$$

$$\frac{\partial u}{\partial x} = \frac{A}{\rho c^2} \frac{\partial p}{\partial t} \quad (2.2.1b)$$

и може да се покаже, че решенията на Уравнения 2.2.1 имат вида

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

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

Първо нека да разгледаме връзката между две съседни тръби.

2.2.1 Преминаване от една тръба в друга

Да разгледаме по-внимателно значението на (2.2.2a).

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_k ще зависи от обемната скорост на вълните, които вървят „напред“ и тази на вълните, които вървят „назад“. За специфична тръба k , Уравнения (2.2.2) ще имат вида:

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

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

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

Вълните, които вървят „напред“ и „назад“, ще означаваме съответно с u^+ и u^- .

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

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

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

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

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

\Leftrightarrow

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

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

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

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

Първо, нека да изразим скоростта 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

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

Заместваме (2.2.7) в (2.2.6)

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

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

$$t_k = \frac{2A_{k+1}}{A_k + A_{k+1}} \quad (2.2.9)$$

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, получена от вълните, които вървят "назад" в тръба $k + 1$. Затова

$$r_k = \frac{A_{k+1} - A_k}{A_k + A_{k+1}} \quad (2.2.10)$$

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

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 скоростта на вълните в тръба k чрез скоростта на вълните в $(k + 1)$ -вата тръба.

Първо разместваме уравнение (2.2.8)

$$u_k^+(t - \tau_k) = u_{k+1}^+(t) \left[\frac{A_k + A_{k+1}}{2A_{k+1}} \right] + u_{k+1}^-(t) \left[\frac{A_k - A_{k+1}}{2A_{k+1}} \right] \quad (2.2.11a)$$

Заместваме (2.2.11a) в (2.2.5)

$$\begin{aligned} u_k^-(t + \tau_k) &= u_k^+(t - \tau_k) - u_{k+1}^+(t) + u_{k+1}^-(t) \\ u_k^-(t + \tau_k) &= u_{k+1}^+(t) \left[\frac{A_k + A_{k+1}}{2A_{k+1}} \right] + u_{k+1}^-(t) \left[\frac{A_k - A_{k+1}}{2A_{k+1}} \right] - u_{k+1}^+(t) + u_{k+1}^-(t) \\ u_k^-(t + \tau_k) &= u_{k+1}^+(t) \left[\frac{A_k + A_{k+1} - 2A_{k+1}}{2A_{k+1}} \right] + u_{k+1}^-(t) \left[\frac{A_k - A_{k+1} + 2A_{k+1}}{2A_{k+1}} \right] \\ u_k^-(t + \tau_k) &= u_{k+1}^+(t) \left[\frac{A_k - A_{k+1}}{2A_{k+1}} \right] + u_{k+1}^-(t) \left[\frac{A_k + A_{k+1}}{2A_{k+1}} \right] \end{aligned} \quad (2.2.11b)$$

Използвайки, че

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

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

можем да запишем Уравнения (2.2.11) във вида:

$$u_k^+(t - \tau_k) = \frac{1}{1 + r_k} u_{k+1}^+(t) - \frac{r_k}{1 + r_k} u_{k+1}^-(t) \quad (2.2.12a)$$

$$u_k^-(t + \tau_k) = -\frac{r_k}{1 + r_k} u_{k+1}^-(t) + u_{k+1}^+(t) + \frac{1}{1 + r_k} u_{k+1}^-(t) \quad (2.2.12b)$$

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 $u_k[t - \tau_k] \xleftrightarrow{\mathcal{FS}} z^{-\tau_k} U_k(z)$

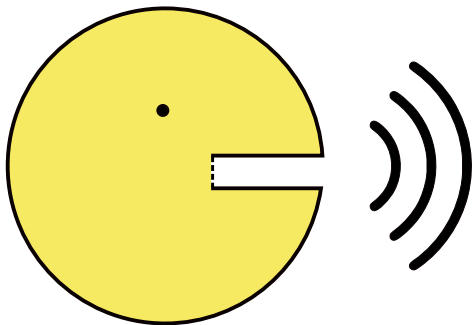
$$\begin{aligned} z^{-\tau_k} U_k^+(z) &= \frac{1}{1+r_k} U_{k+1}^+(z) - \frac{r_k}{1+r_k} U_{k+1}^-(z) \\ z^{\tau_k} U_k^-(z) &= -\frac{r_k}{1+r_k} U_{k+1}^-(z) + U_{k+1}^+(z) + \frac{1}{1+r_k} U_{k+1}^-(z) \\ &\iff \end{aligned} \quad (2.2.13a)$$

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

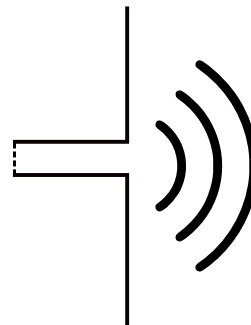
$$U_k^-(z) = -\frac{r_k z^{-\tau_k}}{1+r_k} U_{k+1}^-(z) + U_{k+1}^+(z) + \frac{z^{-\tau_k}}{1+r_k} U_{k+1}^-(z) \quad (2.2.13b)$$

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

В такъв случай може да се покаже, че съществува следната връзка между налягането и обемната скорост:

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

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

$$Z_L(z) = \frac{izL_r R_r}{R_r + izL_r}, \quad (2.2.15)$$

Където $z = e^{i\omega_k}$ описва сигнал с ъглова честота ω_k , L_r и R_r са константи, определени от размера на отвората на устата. За практически цели се избират:

$$R_r = \frac{128}{9\pi^2} > 1 \quad L_r = \frac{8a}{3\pi c}$$

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

При много ниски честоти $Z_L(z) \approx 0$, което значи, че съпротивлението на устните е почти нулево. При средни честоти ($zL_r \ll R_r$), $Z_L(z) \approx izL_r$, а високи честоти, ($zL_r \gg R_r$) $Z_L(z) \approx R_r$. Това значи, че загубите при устните са най-големи при големи честоти, тъй като тогава импедансът е най-голям.

В случая, в който честотата z е висока, $Z_L \approx R_r$ е реално число и не зависи от z , тоест $Z_L(z) = Z_L$.

Товава, ако $p_N(l_N, t) \xleftrightarrow{\mathcal{FS}} P_N(l_N, z)$, $u_N(l_N, t) \xleftrightarrow{\mathcal{FS}} U_N(l_N, z)$ и съответно $Z_L \xleftrightarrow{\mathcal{FS}} Z_L$, 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:

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

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

В случая, в който $Z_L \approx izL_r$ е комплексно, може да се покаже, 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 може да се покаже, че по-добро приближение би било:

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

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

L_G, R_G - константи. Отново предполагайки, че Z_G е реално, тоест честотата ω_k е много ниска, можем да разгледаме уравнение (2.2.17) във времевия домейн.

Нека $u_1(0, t) \xleftrightarrow{\mathcal{FS}} U_1(0, z), p_1(0, t) \xleftrightarrow{\mathcal{FS}} P_1(0, z)$ за фиксиран първи аргумент и съответно $Z_G \xleftrightarrow{\mathcal{FS}} Z_G$:

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

Ако използваме Уравнения (2.2.3), получаваме

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

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

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:

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

Отново може да се покаже, 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:

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

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

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

Тоест дефинираме:

$$U_{N+1}^+(z) = U_L(z) \quad (2.2.21)$$

$$U_{N+1}^-(z) = 0$$

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:

$$U_k = Q_k U_{k+1} \text{ за}$$

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

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

За специалното ограничение глотиса, разгледаме 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

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

Тогава, ако заместим, получаваме:

$$\begin{aligned} \frac{1}{\mathcal{V}(z)} = \frac{U_G(z)}{U_L(z)} &= \frac{\left[\frac{2}{1+r_G}, -\frac{2r_G}{1+r_G} \right] \prod_{i=1}^N Q_i \begin{bmatrix} 1 \\ 0 \end{bmatrix} U_L(z)}{U_L(z)} = \\ &= \left[\frac{2}{1+r_G}, -\frac{2r_G}{1+r_G} \right] \prod_{i=1}^N Q_i \begin{bmatrix} 1 \\ 0 \end{bmatrix} \end{aligned} \quad (2.2.22)$$

Нека изразим Q_k по следния начин:

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

Тогава уравнение (2.2.22) има вида

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

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

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

както е показано в [Пример 3](#)

Може да се покаже, че в общия случай за произволно N , [Уравнение 2.2.23](#) може да се развие итеративно до:

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

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](#).

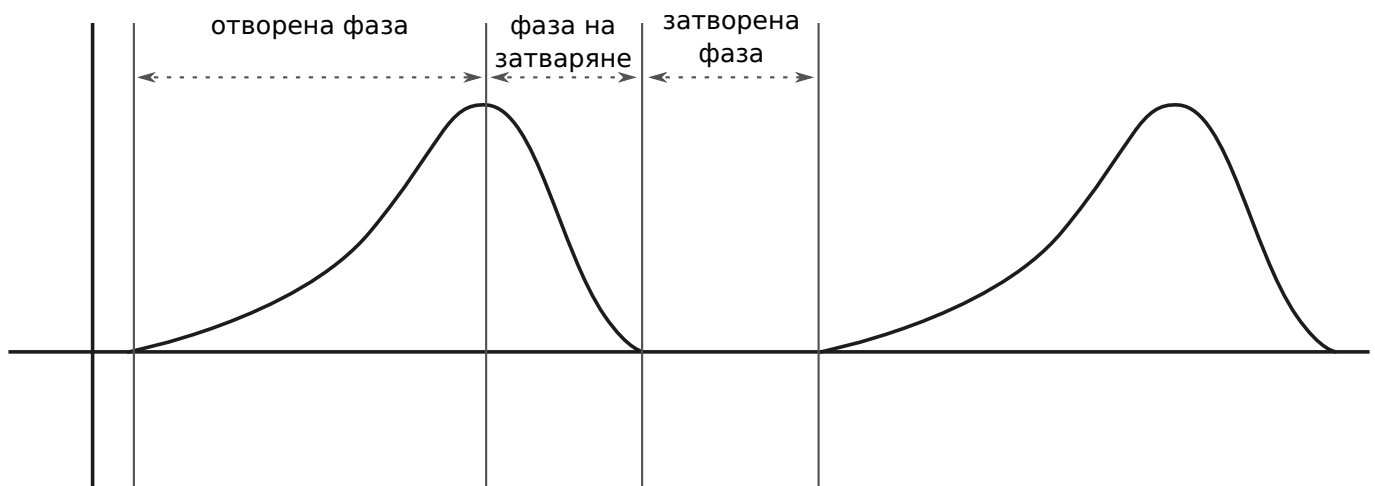
2.2.5 Общ вид на \mathcal{R}

Моделът, описан в [Подраздел 2.2.2](#) се моделира прекалено трудно. Може да се покаже ([\[Тay09\]](#), [\[Quo01\]](#)), че ефектът от радиацията се приближава достатъчно добре с една нула в единичната окръжност, тоест:

$$\mathcal{R}(z) = (1 - \gamma z^{-1}) \quad (2.2.26)$$

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 2.2.3: 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 2.2.3](#). 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.

$$g(t) = \begin{cases} \frac{1}{2}(1 - \cos(\pi n/F_1)), & 0 \leq t \leq F_1 \\ \cos(\pi(n - N_1)/2N_2), & F_1 \leq t \leq F_2 \\ 0, & \text{иначе} \end{cases}$$

Поведението 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 правилно 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

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

Сигналът, показан на [Фигура 2.2.3](#), 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 джитер². Другият ефект, който е важен за истинския човешки глас, е трептенето³, тоест разликата в амплитудите. Освен за естествеността на гласа, тези характеристики могат да носят информация и за емоционалното състояние. Висок джитер може да означава грезгав 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:

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

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:

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

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

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

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 турбуленция, поглъщане, отрязване, 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 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 \leftrightarrow

$$ax_1[n] + bx_2[n] \mapsto ay_1[n] + by_2[n]$$

Дефиниция. (Времево-независима 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-независима.

Свойство 1. Ако системата е времево-независима и сигналът x е перидичен с период N , то и изходът на системата y е перидичен с период N :

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

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

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

Вокалният тракт е времето независима 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-независима 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 характеристикация, 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:

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

Тогаво 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:

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

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

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

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

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

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

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

Да разгледаме фурие преобразуването на [Уравнение 2.3.1](#) за входен сигнал g .

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

$$\frac{y(z)}{\mathcal{G}(z)} = \frac{\sum_{m=0}^M b_m z^{-m}}{\sum_{k=0}^N a_k z^{-k}} \quad (2.3.5)$$

Когато заместим [Уравнение 2.3.5](#) в [Уравнение 2.3.4](#), получаваме

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

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

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

Това означава, че $\mathcal{V}(z)\mathcal{R}(z)$, всъщност описват трансферната 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 $y(z) = \mathcal{G}(z)\mathcal{H}(z)$, 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 $y(z) = \mathcal{G}(z)\mathcal{H}(z)$.

Нека вземем логаритъм от модула:

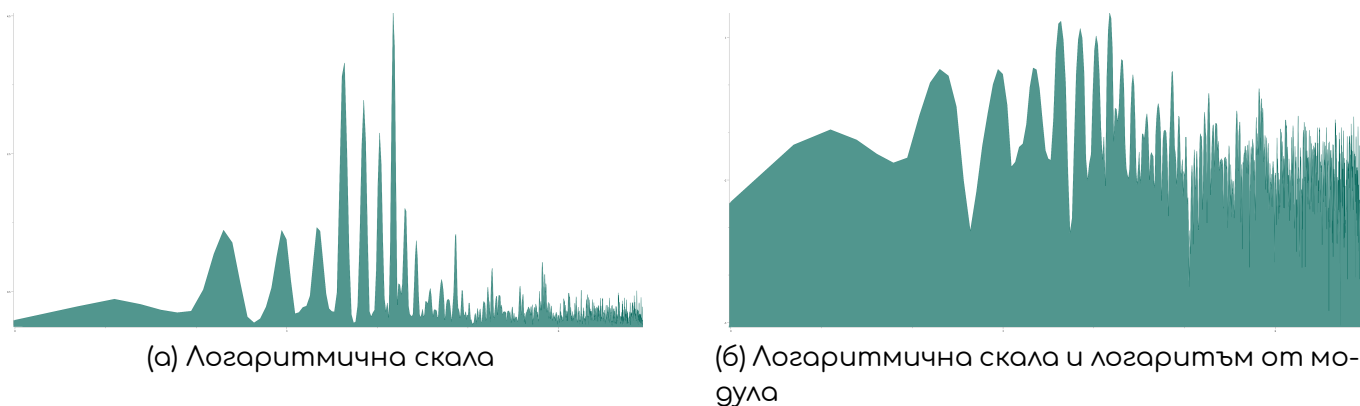
$$\log(|y(z)|) = \log(|\mathcal{G}(z)|) + \log(|\mathcal{H}(z)|)$$

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 вземайки 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).



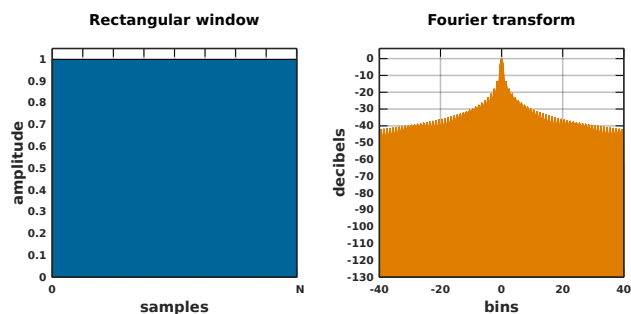
Фигура 2.4.1: Графика на спектъра на сигнал, получен при произнасяне на "а-а-а-а"

Взимането на логаритъм, подчертава периодичността на сигнала, идващ от глотиса. На Фигура 2.4.1а се виждат пикове, породени от фундаменталната честота, на която трепти глотиса, и хармоничните и честоти. Поради загубата на енергията в системата за производство на реч, не всички хармонични честоти имат същата амплитуда. На Фигура 2.4.1б се вижда, че взимането на логаритъм помага за изравняването на хармоничните амплитудати и кара графиката да изглежда "по-периодична". Честотите на Фигура 2.4.1 са изобразени върху логаритмична скала, вместо линейна, тъй като е емпирично установено, че ухото възприема звука логаритмично, но това ще бъде по-подробно разглеждане в следващия подраздел. Сега нека разгледаме логаритмувания спектър като сигнал и му направим Фурие преобразуване до получаване на така нареченият кепстър. Тъй като разглеждаме спектъра 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.

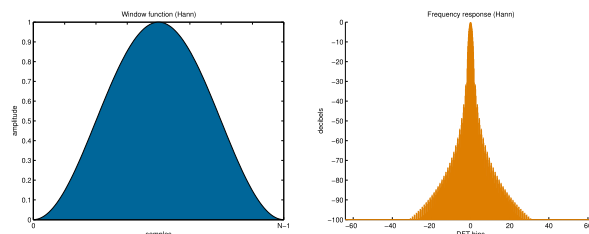
2.4.2 Извличане

Ще извличаме характеристики от подаден аудио файл в wave 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 48Hz. 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 върху две, тоест половината на честотата на дискретизация. Колкото е по-голяма Найкуист честотата, толкова по-добра честотна резолюция получаваме.

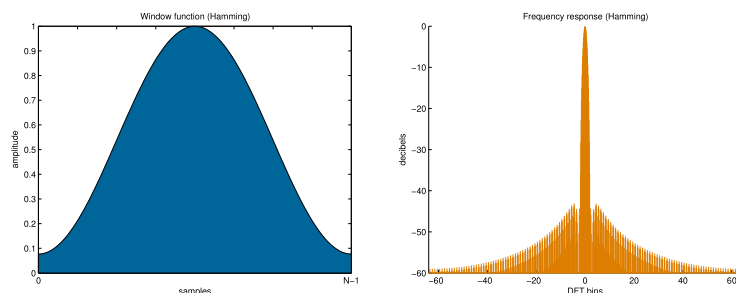
Базирайки се на идеята, че вокалният тракт е статичен за кратък период от време, нахвърляме масива от отделни застъпващи се парчета - фреймове - в рамките на които сигналът е статичен. За да се определи дължината на фрейма, трябва да се вземат предвид две неща: от една страна колкото повече дискрети имаме, толкова по-добра честотна резолюция получаваме. От друга, колкото повече данни взимаме, толкова по-голям е шансът да се смени конфигурации 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. Тази функция е нула навсякъде, освен в избран интервал и обикновено 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.



(а) Правоъгълен прозорец



(б) Прозорец на Хан



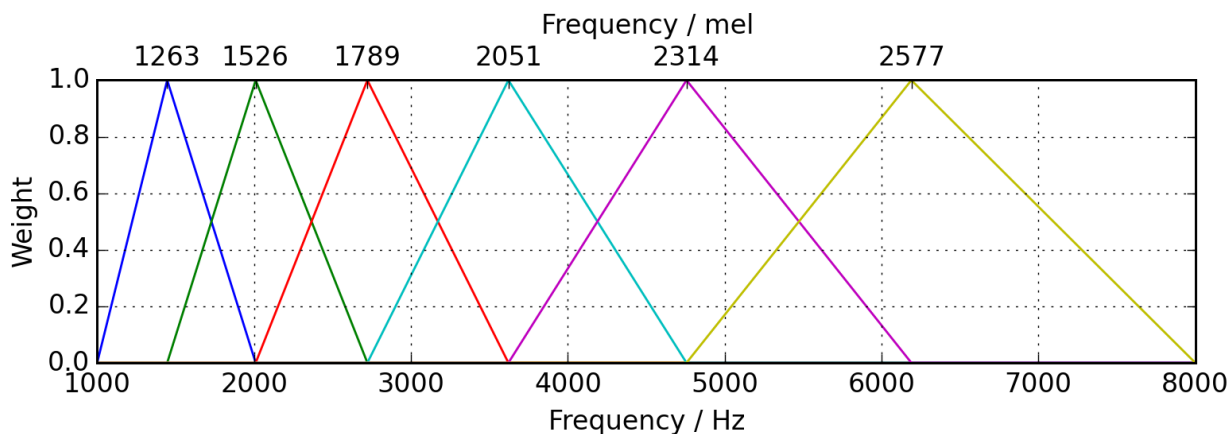
(в) Прозорец на Хеминг

Фигура 2.4.2: Често срещани прозоречни функции

TODO: тука още приказки за прозорци. Прозорецът е моята врата и аз вървя към тях и ги разпитвам.

След като фреймовете вече са периодични, на всеки от тях се прави фурье преобразуване. Може би тука за FFT. За да се подчертае периодичността ~~тадсгфзвезветзвез~~ ГРРРР Чувствителността на човешкото ухо е различна в зависимост от честотата. Всъщност, скалата, по която чуваме звука, е по-скоро логаритмична. Мел скалата е изкуствено (емпрично) създадена, така че разликите между меловите да се усещат като еднакви. Връзката между меловите и честотите е логаритмична, като се задава с формулата:

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



Фигура 2.4.3: Мел

За да се изобрази спектъра в Мел скалата, се натрупват

2.5 Класификация

2.6 Резултати

Глава 3

Грубо в мозъка

3.1 Характеристики

3.1.1 Избор

3.1.2 Извличане

3.2 Класификация

3.3 Резултати

Глава 4

Двойната звезда

4.1 Резултати

Глава 5

Големият портрет

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 определена 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 $\omega_0 = f_0 2\pi$ и се нарича фундаментална ъглова честота.

Товава представянето, което търсим има вида:

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

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

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

Умножаваме [Уравнение A.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

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

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

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

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

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

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

Което означава, че

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

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

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

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:

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

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

$$\begin{aligned} x[n] &= \sum_{k=-\infty}^{\infty} \hat{a}_k e^{\frac{2\pi i k n}{N}} \\ x[n] &= \sum_{k=0}^{N-1} a_k e^{\frac{2\pi i k n}{N}} \end{aligned} \quad (\text{A.1.3})$$

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:

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

и отново използваме, че

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

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

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

A.2 Свойства

- Изместване във времето

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

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

$$\begin{aligned} b_k &= \frac{1}{N} \sum_{n=n_0}^{N-1+n_0} x[n - n_0] e^{-\frac{2\pi i k n}{N}} = \sum_{n=n_0}^{N-1+n_0} x[n - n_0] e^{-\frac{2\pi i k (n-n_0)}{N}} e^{-\frac{2\pi i k n_0}{N}} = \\ e^{-\frac{2\pi i k n_0}{N}} \sum_{\tau=0}^{T-1} x[\tau] e^{-\frac{2\pi i k \tau}{N}} &= e^{-\frac{2\pi i k n_0}{N}} a_k \end{aligned}$$

- Симетричност на комплексно спрегнатите за реален сигнал

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

От уравнение [Уравнение A.1.4](#) следва, че:

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

$$\bar{a}_k = \frac{1}{N} \sum_{n=0}^{N-1} \overline{x[n] e^{-\frac{2\pi i k n}{N}}} = \frac{1}{N} \sum_{n=0}^{N-1} \overline{x[n]} e^{\frac{2\pi i k n}{N}} = \frac{1}{N} \sum_{n=0}^{N-1} x[n] e^{\frac{2\pi i k n}{N}}$$

$$\Rightarrow \bar{a}_k = a_{-k} = a_{N-k}$$

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 $f[n] \xleftrightarrow{\mathcal{FS}} F(e^{i\omega_k})$ и $g[n] \xleftrightarrow{\mathcal{FS}} G(e^{i\omega_k})$ и f, g са периодици с период N

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

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

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

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

Тогава:

$$\begin{aligned}
 H(e^{i\omega_k}) &= c_k = \frac{1}{N} \sum_{n=0}^{N-1} h[n] e^{-\frac{2\pi i k n}{N}} \\
 &= \frac{1}{N} \sum_{n=0}^{N-1} (f * g)[n] e^{-\frac{2\pi i k n}{N}} = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{r=0}^{N-1} f[r] g[n-r] e^{-\frac{2\pi i k n}{N}} \\
 &= \frac{1}{N} \sum_{r=0}^{N-1} f[r] \sum_{n=0}^{N-1} g[n-r] e^{-\frac{2\pi i k n}{N}} \\
 &= \frac{1}{N} \sum_{r=0}^{N-1} f[r] \left[\sum_{n=0}^{N-1} g[n-r] e^{-\frac{2\pi i k n}{N}} \right]
 \end{aligned}$$

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:

$$\begin{aligned}\mathcal{H}(z) &= \frac{\sum_{m=0}^M b_m z^{-m}}{\sum_{k=0}^N a_k z^{-k}} = \\ &= \frac{N(z)}{D(z)} = G \frac{(z - \beta_1)(z - \beta_2) \dots (z - \beta_M)}{(z - \alpha_1)(z - \alpha_2) \dots (z - \alpha_N)},\end{aligned}\tag{2.3.6}$$

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 $\lim_{z \rightarrow \beta_i} \mathcal{H}(z) = 0$

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

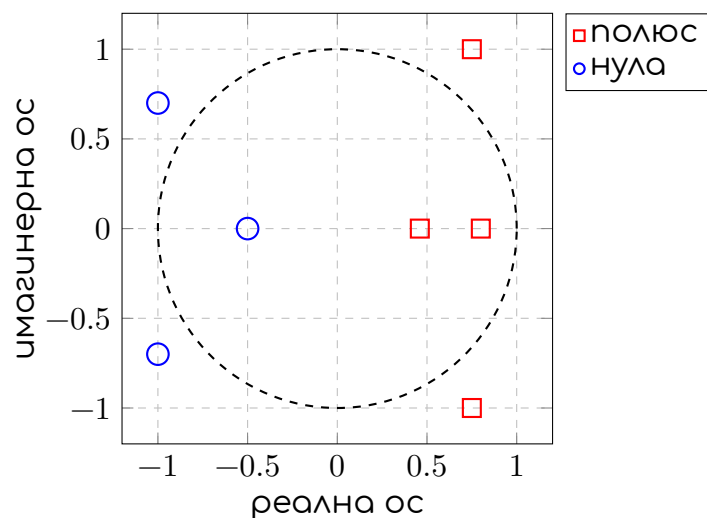
Тъй като коефициентите на $N(z)$ и $D(z)$ са реални, нулите (и съответно полюсите) ще са са или реални, или са част от двойка комплексно спрегнати. Тоест, няма нула (или полюс), която да е комплексна, но да няма комплексно спрегнато из останалите нули (полюси).

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

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

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

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

[illegible]

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 полюси. Следователно, анализирайки тези нули и полюси, можем да направим извод за действието на филтъра.

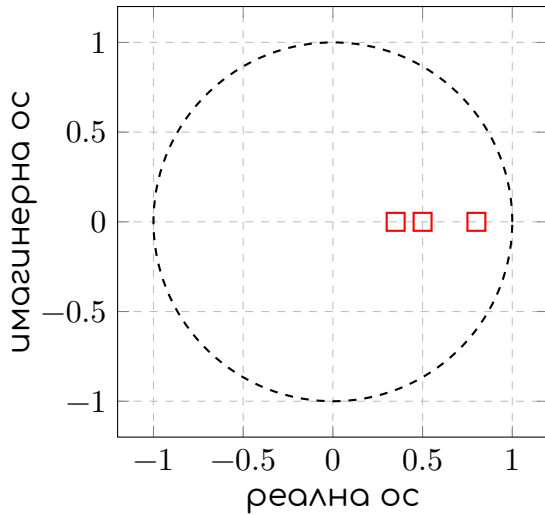
Нека имаме следния филтър от първи ред:

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

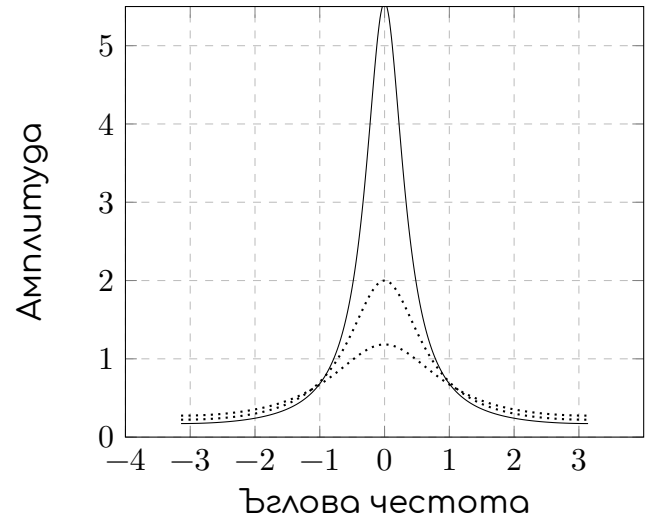
$$\mathcal{H} = G \frac{1}{1 - a_1 z^{-1}},$$

$$G = b_0$$

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

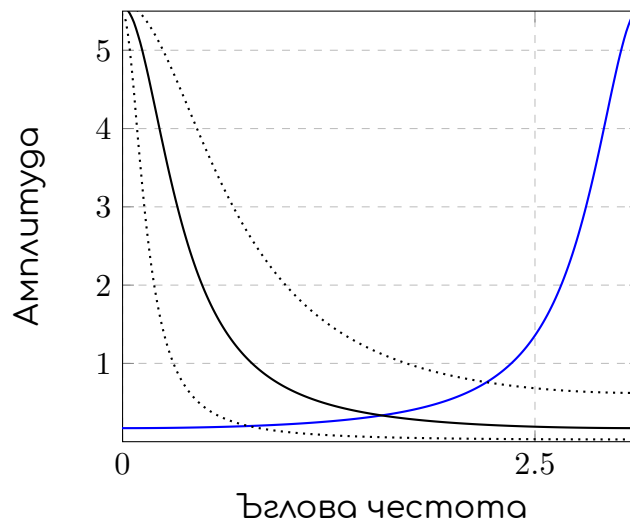


(а) Полус-нула графика за \mathcal{H}



(б) Графика на $g(\omega)$ за $a = 0.7, a = 0.5, a = 0.35$

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

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

$$\begin{aligned} \mathcal{H}(e^{i\omega}) &= \frac{b_0}{1 - a_1 e^{-i\omega}} = \frac{b_0}{1 - a_1 \cos \omega + ia_1 \sin \omega} = \frac{b_0(1 - a_1 \cos \omega - ia_1 \sin \omega)}{1 - 2a_1 \cos \omega + a_1^2} \\ &= \frac{b_0(1 - a_1 \cos \omega)}{1 - 2a_1 \cos \omega + a_1^2} + i \frac{-b_0 a_1 \sin \omega}{1 - 2a_1 \cos \omega + a_1^2} \end{aligned}$$

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 усилят 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](#)).

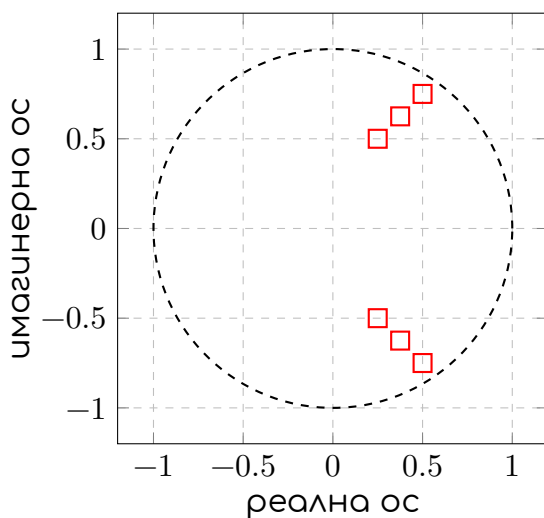
При $a_1 < 0$, филтрите се наричат **високочестотни** (в синьо на [Фигура Б.2.2](#))

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 спряната двойка. Нека разгледаме филтър от втори ред.

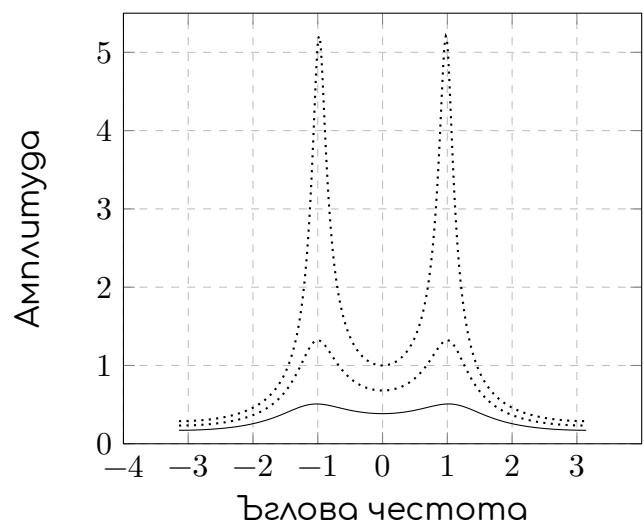
$$y[n] = b_0 x[n] + a_1 y[n-1] + a_2 y[n-2]$$

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

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

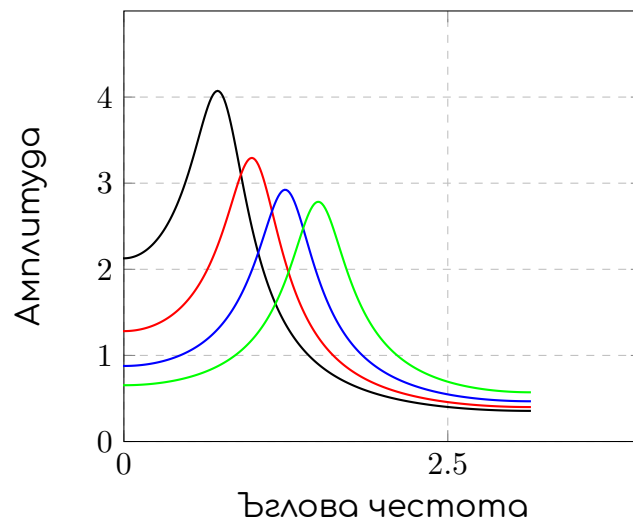


(а) Полус-нула графика за \mathcal{H}

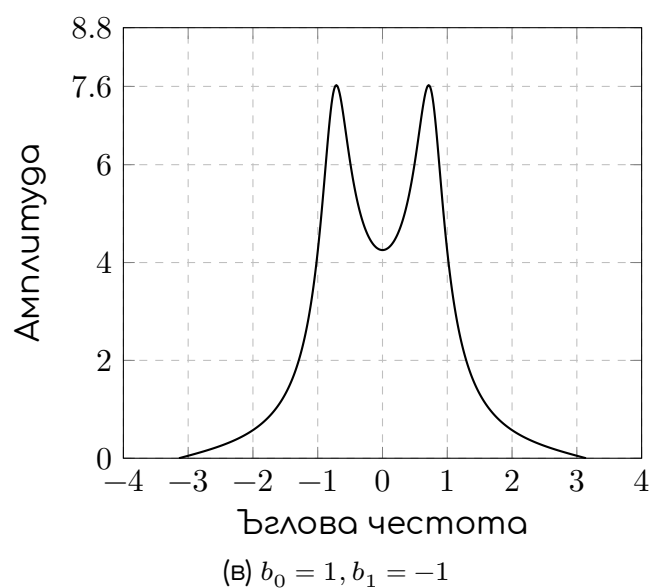
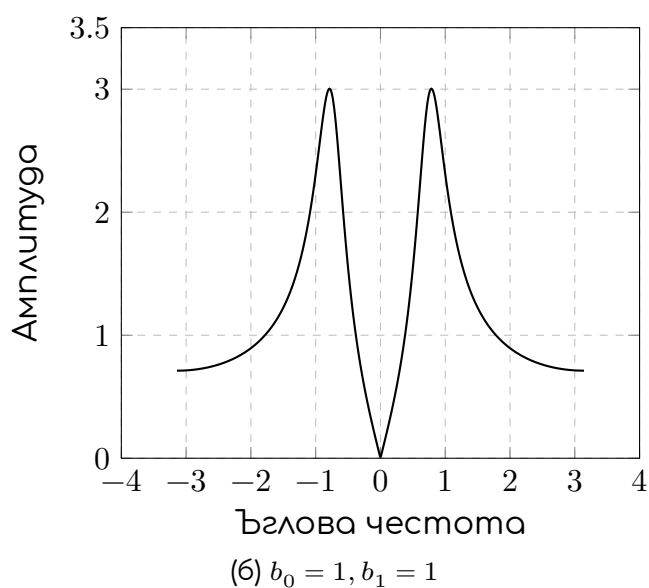
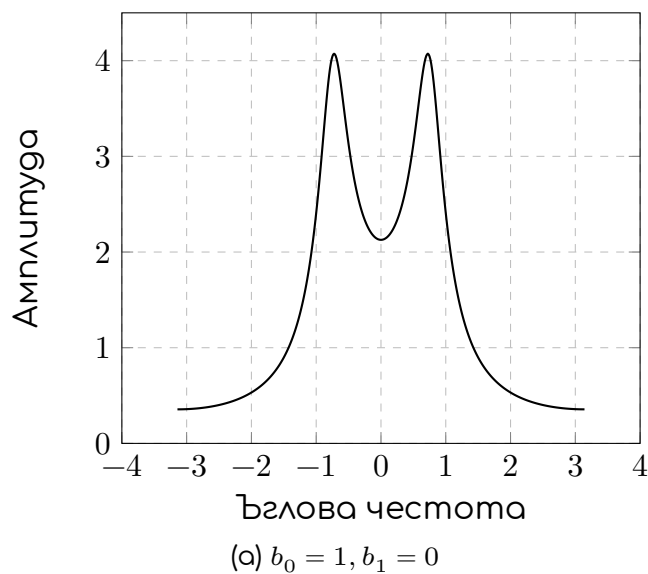


(б) Графика на $g(\omega)$ в $-\pi, \pi$

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



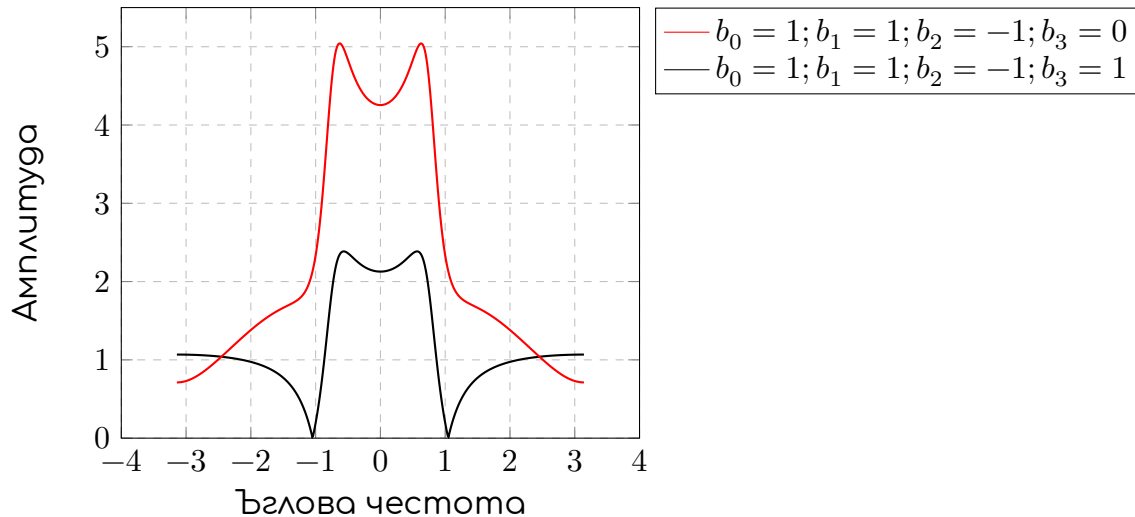
Фигура Б.2.4: Графика на g с отдалечаващи се от реалната ос полюси



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**.

Използвайки вида на филтрите от първи и втори ред, можем да разложим даден сложен филтър \mathcal{H} на произведение от по-прости филтри от първи и втори ред, чиито вид може по-лесно да се моделира чрез промяна на коефициентите. След това съчетаването им е просто произведение в честотния домейн, а свойствата на Фурие преобразуванията ни дават вида и във времевия домейн.

Приложение В

Приложение към Емоции в реч

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

Имаме:

$$U_k = Q_k U_{k+1} \text{ за}$$

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

Търсим: $V(z)$

$$\frac{1}{\mathcal{V}(z)} = \frac{U_G(z)}{U_L(z)} = \left[\frac{2}{1+r_G}, -\frac{2r_G}{1+r_G} \right] \prod_{i=1}^N Q_i \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \quad (2.2.23)$$

$$\begin{aligned}
&= z^{(\tau_1+\tau_2)} \begin{bmatrix} \frac{2}{1+r_G}, & -\frac{2r_G}{1+r_G} \end{bmatrix} \hat{Q}_1 \hat{Q}_2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \\
&= z^{(\tau_1+\tau_2)} \begin{bmatrix} \frac{2}{1+r_G}, & -\frac{2r_G}{1+r_G} \end{bmatrix} \begin{bmatrix} \frac{1}{1+r_1} & \frac{-r_1}{1+r_1} \\ -\frac{r_1 z^{-2\tau_1}}{1+r_1} & \frac{z^{-2\tau_1}}{1+r_1} \end{bmatrix} \begin{bmatrix} \frac{1}{1+r_2} & \frac{-r_2}{1+r_2} \\ -\frac{r_2 z^{-2\tau_2}}{1+r_2} & \frac{z^{-2\tau_2}}{1+r_2} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \\
&= 2z^{(\tau_1+\tau_2)} \begin{bmatrix} \frac{1+r_G r_1 z^{-2\tau_1}}{(1+r_G)(1+r_1)}, & -\frac{r_1+r_G z^{-2\tau_1}}{(1+r_G)(1+r_1)} \end{bmatrix} \begin{bmatrix} \frac{1}{1+r_2} & \frac{-r_2}{1+r_2} \\ -\frac{r_2 z^{-2\tau_2}}{1+r_2} & \frac{z^{-2\tau_2}}{1+r_2} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \\
&= 2z^{(\tau_1+\tau_2)} \begin{bmatrix} \frac{1+r_G r_1 z^{-2\tau_1} + r_1 r_2 z^{-2\tau_2} + r_G r_2 z^{-2(\tau_1+\tau_2)}}{(1+r_G)(1+r_1)(1+r_2)}, & -\frac{r_2+r_G r_1 r_2 z^{-2\tau_1} + r_1 z^{-2\tau_2} + r_G z^{-2(\tau_1+\tau_2)}}{(1+r_G)(1+r_1)(1+r_2)} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
&\iff \\
&\mathcal{V}(z) = \frac{0.5z^{-(\tau_1+\tau_2)}(1+r_G) \prod_{i=1}^2 (1+r_i)}{1+r_G r_1 z^{-2\tau_1} + r_1 r_2 z^{-2\tau_2} + r_G r_2 z^{-2(\tau_1+\tau_2)}}
\end{aligned}$$

□

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

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

Тоест $\tau_1 = \tau_2 = \frac{1}{2}$

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

Заместваме $\tau_i = \frac{1}{2}$ в [Пример 2](#)

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

□

Библиография

- [KNS09] Shashidhar Koolagudi, Sourav Nandy u K Sreenivasa Rao. "Spectral Features for Emotion Classification". B: anp. 2009, c. 1292—1296. DOI: [10.1109/IADCC.2009.4809202](https://doi.org/10.1109/IADCC.2009.4809202).
- [OWN96] Alan V. Oppenheim, Alan S. Willsky u S. Hamid Nawab. Signals & Systems (2Nd Ed.) Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1996. ISBN: 0-13-814757-4.
- [Qua01] Thomas Quatieri -. Discrete-time Speech Signal Processing: Principles and Practice. First. Upper Saddle River, NJ, USA: Prentice Hall Press, 2001. ISBN: 9780132442138.
- [RS78] L. Rabiner u R. Schafer. Digital Processing of Speech Signals. Englewood Cliffs: Prentice Hall, 1978.
- [Tay09] Paul Taylor. Text-to-Speech Synthesis. 1st. New York, NY, USA: Cambridge University Press, 2009. ISBN: 0521899273, 9780521899277.
- [Тал66] Димитър Талев. Гласовете ви чувам. 1966.