

Дипломна работа

Диана Генева <dageneva@qtrp.org>

2018

Съдържание

1	Нулева зона	2
2	Емоции в реч	3
2.1	Физика на тъгата	3
2.2	Приближаване с тръби	5
2.2.1	Преминаване от една тръба в друга	6
2.2.2	Ограничения при устните	10
2.2.3	Ограничения при глотиса	11
2.2.4	Общ вид на \mathcal{V}	11
2.2.5	Общ вид на \mathcal{X}	13
2.2.6	Общ вид на \mathcal{G}	13
2.3	Общ вид на \mathcal{H}	14
2.4	Представяне със системи	15
2.5	Характеристики	17
2.5.1	Избор	17
2.5.2	Извличане	17
2.6	Класификация	17
2.7	Резултати	17
3	Грубо в мозъка	18
3.1	Характеристики	18
3.1.1	Избор	18
3.1.2	Извличане	18
3.2	Класификация	18
3.3	Резултати	18
4	Двойната звезда	19
4.1	Резултати	19
5	Големият портрет	20
A	Приложение за полюси и нули	21
A.1	Дефиниция	21
A.2	Характеризация на филтри	22
B	Приложение към Емоции в реч	28

Глава 1

Нулева зона

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



Глава 2

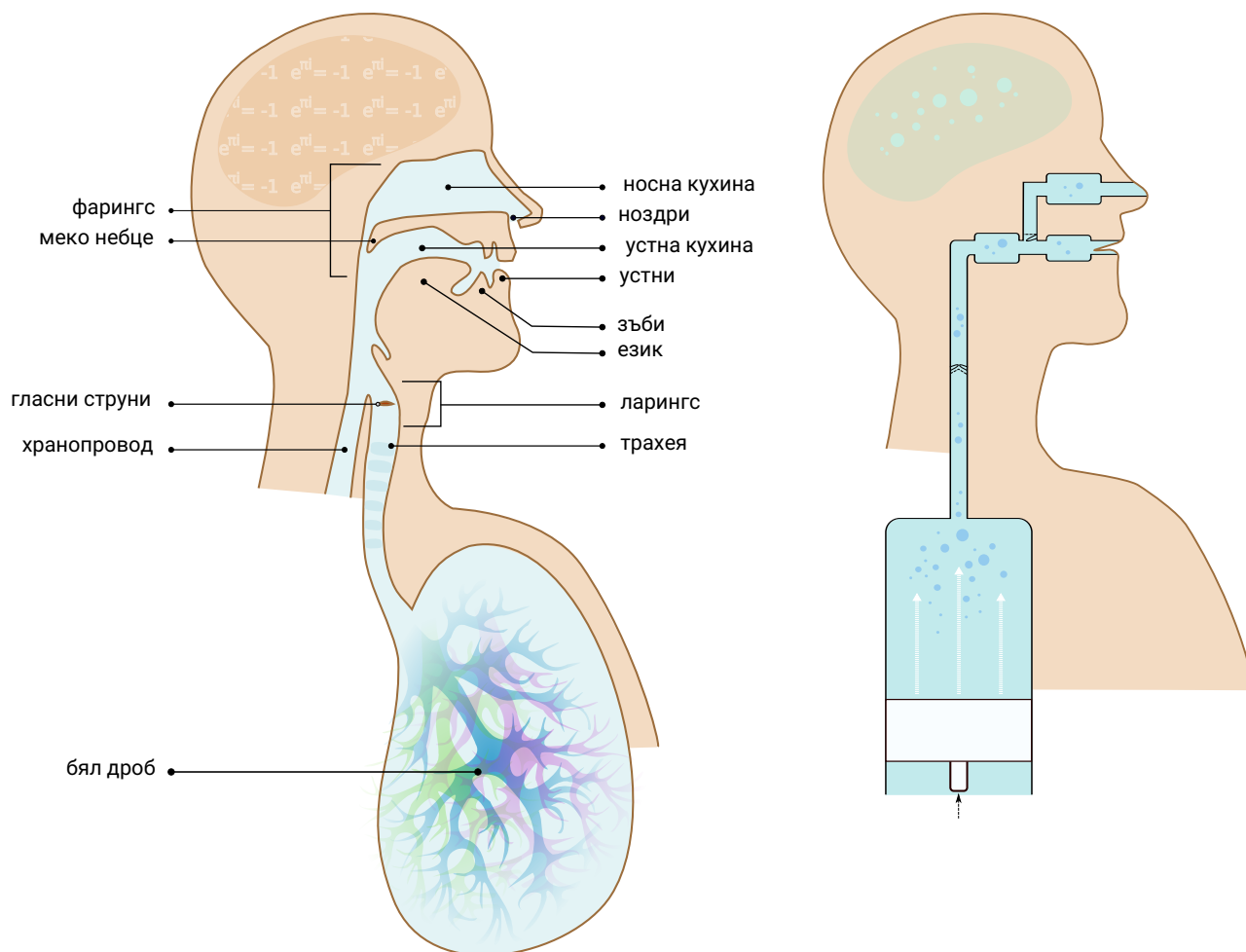
Емоции в реч

2.1 Физика на тъгата

Вокален тракт е общото название на кухините над ларинкса (гръкляна), през които минава въздухът при генериране на реч. При хората той се състои от ларингеална кухина, фаринкс, устна кухина и носна кухина, които могат да се видят на Фиг. 1. Вокалният тракт е отговорен за произвеждане на различни звуци, като текущата конфигурация на отделните му компоненти определя самия звук. Според [<https://ieeexplore.ieee.org/document/4809202>], освен конкретния звук, конфигурацията на вокалния тракт зависи и от емоцията, която изпитва говорещият. Из българската литература се срещат изречения като "страхът стискаше гърлото, задушаваше гласа" [Гласовете ви чувам], а изрази като "буца в гърлото" или "пресъхнало гърло" са станали синоними за дадено емоционално състояние в разговорната реч. Тези фрази имат физиологично обяснение и влияят върху характеристиките на произведената реч. Именно затова бихме искали да можем математически да моделираме конфигурацията на вокалния тракт и да извличаме характеристики от нея.

Да разгледаме по-подробно 2.1.1 и цялостна система за производство на реч. Речта, всъщност, представлява просто акустичната вълна, получена на края на системата - устни и ноздри - в следствие на изкарания 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 пропъгва нагоре по трахеята и през глотиса (отворе между гласните струни). Тъй като налягането в глотиса е по-малко от това в който и да е от двата му края, по закона на Бернули в някакъв момент става толкова ниско, че позволява на гласните струни да се затворят. В следствие се натрупва налягане зад гласните струни, 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 въл-



(а) Физическо представяне на вокалния тракт

(б) Опростено представяне на вокалния тракт

Фигура 2.1.1: Система за производство на реч

на, която се получава в следствие на осцилацията, преминава през вокалния тракт, където се завихря при срещане на презгради 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: стените на вокалния тракт са меки и еластични, радиацията на вълната при устните и ноздрите и други.

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 презграда по пътя си (като например зъби, устни) и се получава турбуленция при опита въздухът да бъде избутан през презградата
3. Презградни (експлозивни) Те се получават, когато презградата е пълна, което позволява да се натрупа налягане зад съответната презгада, и това налягане се освобождава рязко.

Ако разгледаме вокалния тракт и носната кухина като свързани тръби с

постоянно напречно сечение и вземем предвид горното описание за генериране на звук, то честотният спектър ще зависи от честотната пропускливост на тръбите (frequency selectivity) Трябва да видя онова в тетрадката. Това много прилича на свирене на духов инструмент. Честотите, на които се получава резонанс, зависят от формата и размера на тръбите. Известно е, че за да се образува определен звук, трябва да се промени формата на вокалния тракт по съответния начин. Например, когато човек казва "н", езикът се залепа зад зъбите и т.н.. В такъв случай, спектралните особености на сигнала се менят с времето, тъй като се мени и положението на вокалния тракт. Смята се [], че състоянието на вокалния тракт е сравнително статично (достатъчно статично за нашите цели) в рамките на 15ms, преди да се смени съответната фонема, която се изговаря. В такъв случай, можем да изследваме спектралните свойства в този отрязък от време и да извличаме информацията за подлежащата емоция в него.

2.2 Приближаване с тръби

За улеснение, нека разгледаме конкретна конфигурация. Например тази, при произнасянето на фонемата "ъ", тъй като е възможно най-проста.

В случая на гласната "ъ", глотисът трепти, а клапата към носната кухина е затворена и само устната кухина участва в образуването на звука. Т.е. имаме някакъв източник на псевдо-периодично трептене g - глотис, после вълната преминава и се променя от вокалния тракт v и накрая излиза и се пречупва през устните r . Сигналът, който получаваме накрая, е резултат от прилагането на филтъра на вокалния тракт и устните върху сигнала от глотиса. Това може да се запише като конволюция във времевия домейн. $y(t) = g(t)*v(t)*r(t)$ или произведение в честотния $Y(z) = G(z)V(z)R(z)$, където

$$G(z) = \mathcal{Z}\{g(t)\}$$

$$V(z) = \mathcal{Z}\{v(t)\}$$

$$R(z) = \mathcal{Z}\{r(t)\}$$

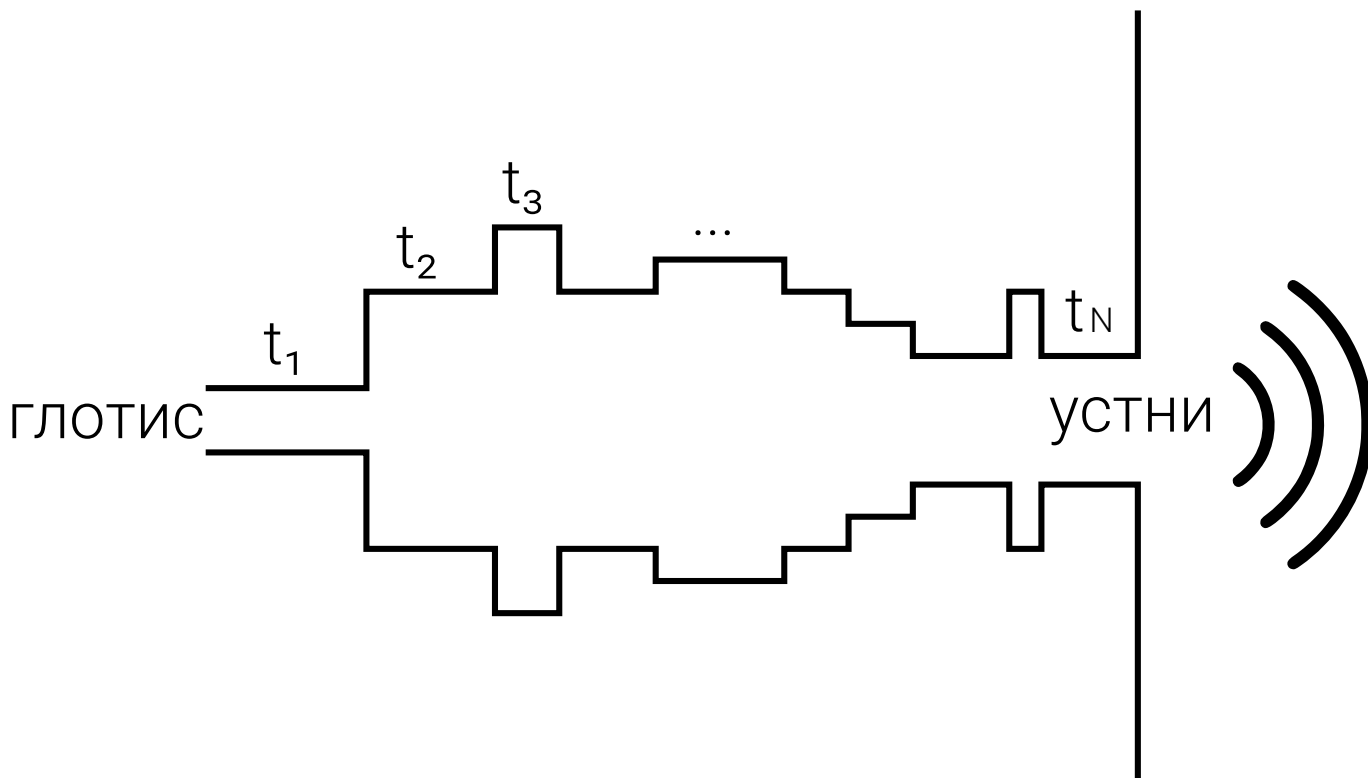
$$Y(z) = \mathcal{Z}\{y(t)\}$$

Това, от което се интересуваме, е \mathcal{V} .

Нека устната кухина е приближена с N на брой тръби с постоянно напречно сечение, както е показано на Фигура 2.2.1 За още по-голямо опростяване, нека няма и загуба на енергия, каквата би се получила по принцип.

Нека въведем следните стандартни означения:

1. c - скорост на звука, т.е. това е разстоянието, което пропътува вълната за единица време
2. ρ - плътност на въздуха
3. A - напречно сечение
4. u - обемна скорост



Фигура 2.2.1: Приближение на вокалния тракт с N тръби

5. p - налягане - можем да си мислим за него като “колко сила е нужна, за да се преместят частици”

Нека означим с u_L обемната скорост в последната точка на последната тръба t_N , тоест $u_L(t) = u_N(l_N, t)$, тъй като тя отговаря за изхода при усните, а с u_G обемната скорост съответстваща първата точка на t_1 , тоест $u_G(t) = u_1(0, t)$, макар това да е изпълнено само в идеалния случай. Тогава можем да опишем отношението на u_L и u_G по следния начин: обемната скорост при усните е резултат от прилагането на филтъра на вокалния тракт върху сигнала от глотиса $u_L(t) = u_G(t) * v(t) \leftrightarrow v(t) = \frac{u_L}{u_G}$. Или изразено в честотния домейн като:

$$\mathcal{V} = \frac{\mathcal{U}_L}{\mathcal{U}_G}$$

Първо да разгледаме връзката между две съседни тръби.

2.2.1 Преминаване от една тръба в друга

Тъй като въздухът е течна среда, как пътува вълна през него се определя от свойствата на въздуха и от ограниченията на тръбата като дължина и ширина. Звуковите вълни, преминаващи през тръба, изпълняват уравнения [Rabiner and Schafer, pg 59, (3.2a) (3.2b)]:

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

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

При преминаване от една тръба в друга, част от вълните ще преминат към следващата тръба, а част от тях ще се отразят в наобратно. В такъв случай във всеки момент от време t и във всяка точка x на всяка тръба, обемната скорост u ще зависи от обемната скорост на вълните, които вървят "напред" и тази на вълните, които вървят "назад". Затова може да се покаже, че уравненията (2.2.1) имат решение във следния вид:

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

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

където

1. t_k е k -тата тръба, която има дължина l_k и лице на напречното сечение A_k
2. x е разстояние в t_k ($0 \leq x \leq l_k$)
3. t е момент от време
4. u^+ и u^- са съответно вълните, които вървят "напред" и тези, които вървят "назад"

Освен това, при преминаването от една тръба в друга са изпълнени следните равенства:

$$u_k(l_k, t) = u_{k+1}(0, t) \quad (2.2.3a)$$

$$p_k(l_k, t) = p_{k+1}(0, t) \quad (2.2.3b)$$

Когато заместим уравнения (2.2.3) в (2.2.2), 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:

$$u_k^+(t - \tau_k) - u_k^-(t + \tau_k) = u_{k+1}^+(t) - u_{k+1}^-(t) \quad (2.2.4)$$

$$\frac{A_{k+1}}{A_k} [u_k^+(t - \tau_k) + u_k^-(t + \tau_k)] = u_{k+1}^+(t) + u_{k+1}^-(t) \quad (2.2.5)$$

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 (2.2.4) получаваме

$$u_k^-(t + \tau_k) = u_k^+(t - \tau_k) - u_{k+1}^+(t) + u_{k+1}^-(t) \quad (2.2.6)$$

Заместваме (2.2.6) в (2.2.5)

$$\begin{aligned} u_{k+1}^+(t) &= \frac{A_{k+1}}{A_k} u_k^+(t - \tau_k) + \frac{A_{k+1}}{A_k} [u_k^+(t - \tau_k) - u_{k+1}^+(t) + u_{k+1}^-(t)] - u_{k+1}^-(t) \\ u_{k+1}^+(t) \left[1 + \frac{A_{k+1}}{A_k} \right] &= u_k^+(t - \tau_k) \frac{2A_{k+1}}{A_k} + u_{k+1}^-(t) \left[\frac{A_{k+1}}{A_k} - 1 \right] \\ u_{k+1}^+(t) \left[\frac{A_k + A_{k+1}}{A_k} \right] &= u_k^+(t - \tau_k) \frac{2A_{k+1}}{A_k} + u_{k+1}^-(t) \left[\frac{A_{k+1} - A_k}{A_k} \right] \\ u_{k+1}^+(t) &= u_k^+(t - \tau_k) \left[\frac{2A_{k+1}}{A_k + A_{k+1}} \right] + u_{k+1}^-(t) \left[\frac{A_{k+1} - A_k}{A_k + A_{k+1}} \right] \end{aligned} \quad (2.2.7)$$

Коефициентът, който стои пред $u_k^+(t - \tau_k)$ в уравнение (2.2.7), представлява количеството енергия, която преминава от тръба k в тръба $k + 1$, идваща от вълните, които се движат "напред" в k -тата тръба. Затова

$$t_k = \frac{2A_{k+1}}{A_k + A_{k+1}} \quad (2.2.8)$$

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

$$r_k = \frac{A_{k+1} - A_k}{A_k + A_{k+1}} \quad (2.2.9)$$

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 (2.2.9), $r_k = 0$, а от (2.2.8) 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 (2.2.7)

$$u_k^+(t - \tau_k) = u_{k+1}^+(t) \left[\frac{A_k + A_{k+1}}{2A_{k+1}} \right] + u_{k+1}^-(t) \left[\frac{A_k - A_{k+1}}{2A_{k+1}} \right] \quad (2.2.10a)$$

Заместваме (2.2.10a) в (2.2.4)

$$\begin{aligned} u_k^-(t + \tau_k) &= u_k^+(t - \tau_k) - u_{k+1}^+(t) + u_{k+1}^-(t) \\ u_k^-(t + \tau_k) &= u_{k+1}^+(t) \left[\frac{A_k + A_{k+1}}{2A_{k+1}} \right] + u_{k+1}^-(t) \left[\frac{A_k - A_{k+1}}{2A_{k+1}} \right] - u_{k+1}^+(t) + u_{k+1}^-(t) \\ u_k^-(t + \tau_k) &= u_{k+1}^+(t) \left[\frac{A_k + A_{k+1} - 2A_{k+1}}{2A_{k+1}} \right] + u_{k+1}^-(t) \left[\frac{A_k - A_{k+1} + 2A_{k+1}}{2A_{k+1}} \right] \\ u_k^-(t + \tau_k) &= u_{k+1}^+(t) \left[\frac{A_k - A_{k+1}}{2A_{k+1}} \right] + u_{k+1}^-(t) \left[\frac{A_k + A_{k+1}}{2A_{k+1}} \right] \end{aligned} \quad (2.2.10b)$$

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 уравненията (2.2.10) във вида:

$$u_k^+(t - \tau_k) = \frac{1}{1 + r_k} u_{k+1}^+(t) - \frac{r_k}{1 + r_k} u_{k+1}^-(t) \quad (2.2.11a)$$

$$u_k^-(t + \tau_k) = -\frac{r_k}{1 + r_k} u_{k+1}^-(t) + u_{k+1}^+(t) + \frac{1}{1 + r_k} u_{k+1}^-(t) \quad (2.2.11b)$$

Сега да разгледаме уравнения (2.2.11) в z -домейна:

$$\begin{aligned} z^{-\tau_k} U_k^+(z) &= \frac{1}{1 + r_k} U_{k+1}^+(z) - \frac{r_k}{1 + r_k} U_{k+1}^-(z) \\ z^{\tau_k} U_k^-(z) &= -\frac{r_k}{1 + r_k} U_{k+1}^-(z) + U_{k+1}^+(z) + \frac{1}{1 + r_k} U_{k+1}^-(z) \\ U_k^+(z) &= \frac{z^{\tau_k}}{1 + r_k} U_{k+1}^+(z) - \frac{r_k z^{\tau_k}}{1 + r_k} U_{k+1}^-(z) \end{aligned} \quad (2.2.12a)$$

$$U_k^-(z) = -\frac{r_k z^{-\tau_k}}{1 + r_k} U_{k+1}^-(z) + U_{k+1}^+(z) + \frac{z^{-\tau_k}}{1 + r_k} U_{k+1}^-(z) \quad (2.2.12b)$$

2.2.2 Ограничения при устните

Както се вижда от фигура Фигура 2.2.1, изходът при устните представлява отвор, който пречупва радиращите звуковите вълни. За да се опишат ефектите от това пречупване, е нужно да знаем връзката между налягането и обемната скорост. В реалния случай това би било трудно, затова за улеснение ще приемем, че изходът при устните е презграден от безкрайна равнина. В такъв случай може да се покаже, че е изпълнено:

$$\mathcal{P}_L(z) = Z_L(z)\mathcal{U}_L(z), \quad (2.2.13)$$

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

$$Z_L(z) = \frac{jzL_rR_r}{R_r + jzL_r},$$

Където z е честота, L_r и R_r са константи, определени от размера на отвора на устата. За практически цели се избират:

$$R_r = \frac{128}{9\pi^2} > 1$$

$$L_r = \frac{8a}{3\pi c}$$

a - радиус на отвора, 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 ($zL_r \ll R_r$), $Z_L(z) \approx jzL_r$, 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 . Тогава можем да разгледаме уравнението (2.2.13) във времевия домейн, ще получим:

$$p_L(t) = Z_L u_L(t)$$

$$\frac{\rho c}{A_L} [u_N^+(t - \tau_N) + u_N^-(t + \tau_N)] = Z_L [u_N^+(t - \tau_N) - u_N^-(t + \tau_N)]$$

$$u_N^-(t + \tau_N) \frac{(\rho c + A_L Z_L)}{A_L} = u_N^+(t - \tau_N) \frac{(A_L Z_L - \rho c)}{A_L}$$

$$u_N^-(t + \tau_N) = -r_L u_N^+(t - \tau_N), \text{ където}$$

$$r_L = \left(\frac{\rho c - A_L Z_L}{\rho c + A_L Z_L} \right) \quad (2.2.14)$$

Като заместим в уравнение (2.2.16), получаваме

$$\begin{aligned} u_N(l_N, t) &= u_N^+(t - \tau_N) - u_N^-(t + \tau_N) = \\ &= (1 + r_L) u_N^+(t - \tau_N) \end{aligned} \quad (2.2.15)$$

В случая, в който $Z_L \approx jzL_r$ е комплексно, може да се покаже, че уравнение (2.2.14) 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, трябва да се отчете фактът, че когато глотисът е затворен, импедансът е безкраен, а обемната скорост нулева. Връзката $U_1(0, z) = U_G(z)$, която използвахме досега, е твърде наивна, затова едно по-добро приближение би било:

$$U(0, z) = U_G(z) - \frac{P(0, z)}{Z_G(z)}, \quad (2.2.16)$$

където Z_G описва импеданса на глотиса и $Z_G(z) = R_G + jzL_G$,

L_G, R_G —константи. Отново предполагайки, че Z_G е реално, тоест честотата z е много ниска, можем да разгледаме уравнение (2.2.16) във времевия домейн:

$$\begin{aligned} u_1^+(t) - u_1^-(t) &= u_G(t) - \frac{\rho c}{A_1} \left[\frac{u_1^+(t) + u_1^-(t)}{Z_G} \right] \\ u_1^+(t) \left[1 + \frac{\rho c}{A_1 Z_G} \right] &= u_G(t) + u_1^-(t) \left[1 - \frac{\rho c}{A_1 Z_G} \right] \\ u_1^+(t) &= u_G(t) \left[\frac{A_1 Z_G}{A_1 Z_G + \rho c} \right] + u_1^-(t) \left[\frac{A_1 Z_G - \rho c}{A_1 Z_G + \rho c} \right] \\ u_1^+(t) &= u_G(t) \left[\frac{1 + r_G}{2} \right] + r_G u_1^-(t) \\ U_G(z) &= \left[\frac{2}{1 + r_G} \right] U_1^+(z) - \left[\frac{2r_G}{1 + r_G} \right] U_1^-(z), \text{ където} \end{aligned} \quad (2.2.17)$$

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

и използвайки, че

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

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 всички тръби имат равна дължина и $\tau_i = \frac{1}{2}, i \in [1...N]$ Тогава уравнения (2.2.12) имат вида:

$$U_k^+(z) = \frac{z^{1/2}}{1 + r_k} U_{k+1}^+(z) - \frac{r_k z^{1/2}}{1 + r_k} U_{k+1}^-(z) \quad (2.2.18a)$$

$$U_k^-(z) = -\frac{r_k z^{-1/2}}{1 + r_k} U_{k+1}^-(z) + U_{k+1}^+(z) + \frac{z^{-1/2}}{1 + r_k} U_{k+1}^-(z) \quad (2.2.18b)$$

За да опишем граничните условия при устните, ще приемем, че $U_{N+1}(z)$ е z -трансформацията на входа на $(N + 1)$ -та тръба, която е безкрайно дълга, и заради това скоростта на вървящите "назад" вълни трябва да бъде 0. Тоест,

$$\begin{aligned} U_{N+1}^+(z) &= U_L(z) \\ U_{N+1}^-(z) &= 0 \end{aligned} \quad (2.2.19)$$

Ако представим уравнения (2.2.18) в матричен вид, получаваме:

$$U_k = Q_k U_{k+1} \text{ за}$$

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

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

Сега представяме уравнение (2.2.17) като

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

Ограниченията за (2.2.19) 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

$$\frac{1}{\mathcal{V}(z)} = \frac{U_G(z)}{U_L(z)} = \begin{bmatrix} \frac{2}{1+r_G} & -\frac{2r_G}{1+r_G} \end{bmatrix} \prod_{i=1}^N Q_i \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (2.2.20)$$

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 (2.2.20) има вида

$$\frac{1}{\mathcal{V}(z)} = \frac{U_G(z)}{U_L(z)} = z^{N/2} \left[\frac{2}{1+r_G}, -\frac{2r_G}{1+r_G} \right] \prod_{i=1}^N \hat{Q}_i \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad (2.2.21)$$

Нека изразим общия вид на $\mathcal{V}(z)$. За $N = 2$, например, (2.2.21) има вида:

$$\mathcal{V}(z) = \frac{0.5(1+r_G) \prod_{i=1}^2 (1+r_i) z^{-1}}{1 + (r_G r_1 + r_1 r_2) z^{-1} + (r_G r_2) z^{-2}}, \quad (2.2.22)$$

както е показано в Пример 1

За всяка стойност на N в числител се получава произведение $(1+r_G) \prod_{i=1}^N (1+r_i)$, а в знаменател се получава полином на z и от основната теорема на алгебрата следва, че уравнението (2.2.21) ще има решения от вида:

$$\mathcal{V}(z) = \frac{0.5(1+r_G) \prod_{i=1}^N (1+r_i) z^{-N/2}}{1 - \sum_{i=1}^N \alpha_i z^{-i}} \quad (2.2.23)$$

2.2.5 Общ вид на \mathcal{R}

За да моделираме устните, трябва да се отрази радиацията. Искаме за радиацията да е изпълнено Уравнение 2.2.12, като

$$\mathcal{R}(z) = P_L(z) U_L(z)$$

Също така видяхме, че налягането и скоростта са свързани с високочестотен филтър, следователно може да се опише от КИХ филтър, тоест:

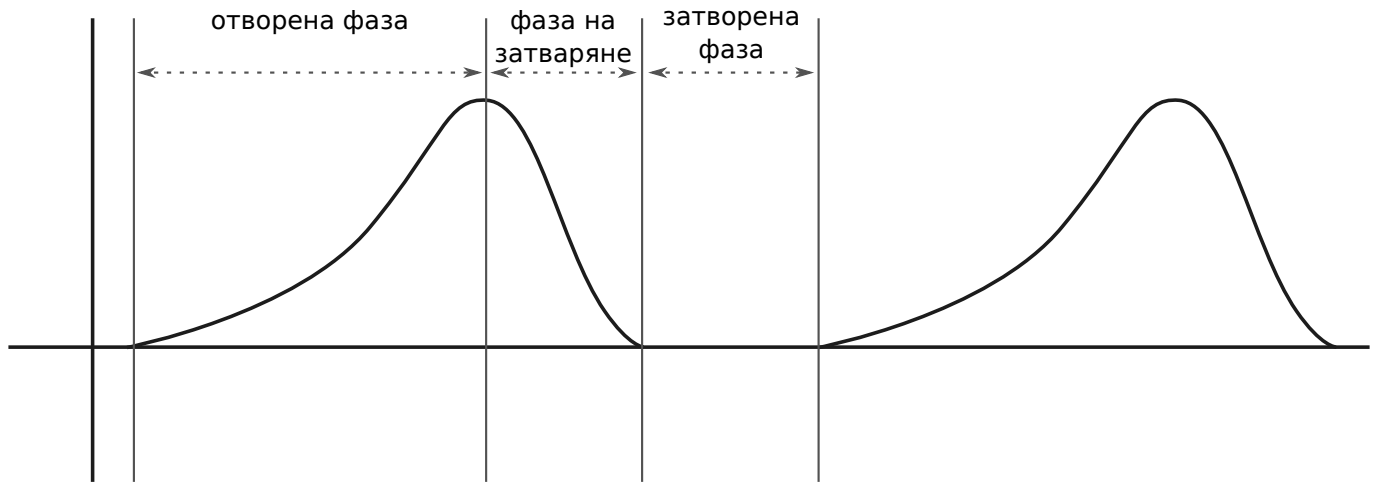
$$\mathcal{R}(z) = (1 - \gamma z^{-1}) \quad (2.2.24)$$

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. При звучните, както описахме по-горе, глотисът произвежда квази-периодичен сигнал, а при беззвучните се получава случаен шум.

Да разгледаме вида на \mathcal{G} в случая на озвучен звук.

На Фигура 2.2.2 се вижда, че импулсът от глотиса може да се раздели на три фази:



Фигура 2.2.2: Пример за импулс от глотиса

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

Съответно ще моделираме функцията g в зависимост от фазата. По-лесно ще е да се направи във времевия домейн. Едно добро приближение е:

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

Тъй като функцията g има краен брой точки, в които не е нула, z -трансформацията ѝ ще има две нули. Тоест може да се моделира като:

$$\mathcal{G}(z) = (1 - \beta_1 z^{-1})(1 - \beta_2 z^{-1}) \quad (2.2.25)$$

В случая на беззвучие за \mathcal{G} се използва някакъв генератор на случайни числа, за да се симулира шум.

2.3 Общ вид на \mathcal{H}

От уравнения 2.2.22, 2.2.24, 2.2.25 следва, че видът на \mathcal{H} е

$$y(z) = \mathcal{G}(z)\mathcal{V}(z)\mathcal{R}(z) = (1 - \beta_1 z^{-1})(1 - \beta_2 z^{-1})(1 - \gamma z^{-1}) \frac{0.5(1 + r_G) \prod_{i=1}^N (1 + r_i) z^{-N/2}}{1 - \sum_{i=1}^N \alpha_i z^{-i}} \quad (2.3.1)$$

2.4 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.

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

$$\sum_{k=0}^N a_k y[n - k] = \sum_{m=0}^M b_m x[n - m] \quad (2.4.1)$$

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 вокалния тракт е линейна, времево-независима система, която удовлетворява уравнение Уравнение 2.4.1, 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 характеристика, 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 дискретният 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:

$$g[n] = \sum_{k=-\infty}^{\infty} g[k]\delta[n-k] \mapsto \sum_{k=-\infty}^{\infty} g[k]h_k[n] = y[n] \quad , \quad (2.4.2)$$

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 2.4.2 има вида:

$$y[n] = \sum_{k=-\infty}^{\infty} g[k]h_k[n] = \sum_{k=-\infty}^{\infty} g[k]h[n-k] \quad , \quad (2.4.3)$$

Дефиниция. Дискретна конволюция

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

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

Тоест $y[n] = (g * h)[n]$ Ако запишем това равенство в z -домейна, получаваме

$$Y(z) = G(z)H(z)$$

$$H(z) = \frac{Y(z)}{G(z)} \quad (2.4.4)$$

H се нарича предавателна функция за системата.

Да разгледаме Уравнение 2.4.1 в z -домейна.

$$\begin{aligned} \left[\sum_{k=0}^N a_k z^{-k} \right] Y(z) &= \left[\sum_{m=0}^M b_m z^{-m} \right] G(z) \\ \frac{Y(z)}{G(z)} &= \frac{\sum_{m=0}^M b_m z^{-m}}{\sum_{k=0}^N a_k z^{-k}} \end{aligned} \quad (2.4.5)$$

Когато заместим Уравнение 2.4.5 в Уравнение 2.4.4, получаваме

$$H(z) = \frac{\sum_{m=0}^M b_m z^{-m}}{\sum_{k=0}^N a_k z^{-k}} \quad (2.4.6)$$

2.5 Характеристики

2.5.1 Избор

2.5.2 Извличане

Първо се изчита wav файла, като данни се запазват в масив от float числа. Базирайки се на идеята, че вокалния тракт е статичен за много къс период от време, нахвърляме масива с данните на отделни застъпващи се фреймове, в рамките на които сигнала е статичен ("представяме си, че е статичен"). За да получим добра честотна резолюция, трябва да се включат голям брой samples от файла, но колкото по-голяма е дължината на фрейма, толкова по-голям е шансът да включим данни за различни конфигурации на гласа. За да се постигне някакъв trade-off между двете, обикновено стойностите, които се избират, са 25 милисекунди за дължина на фрейм и 10 милисекунди за разстояние между два последователни фрейма.[Paul]

Тъй като алгоритъмът за извличане на mfcc коефициенти, изисква броя на сеплите да е степен на 2, допълваме последните фреймове с 0, ако това е нужно, тъй като това не влияе на точността. [мжое би онази статия за fourier]

2.6 Класификация

2.7 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

Приложение за полюси и нули

А.1 Дефиниция

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

$$\mathcal{H}(z) = \frac{\sum_{m=0}^M b_m z^{-m}}{\sum_{k=0}^N a_k z^{-k}} = \quad (2.4.6)$$

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

където $G = b_M/a_N$ 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).

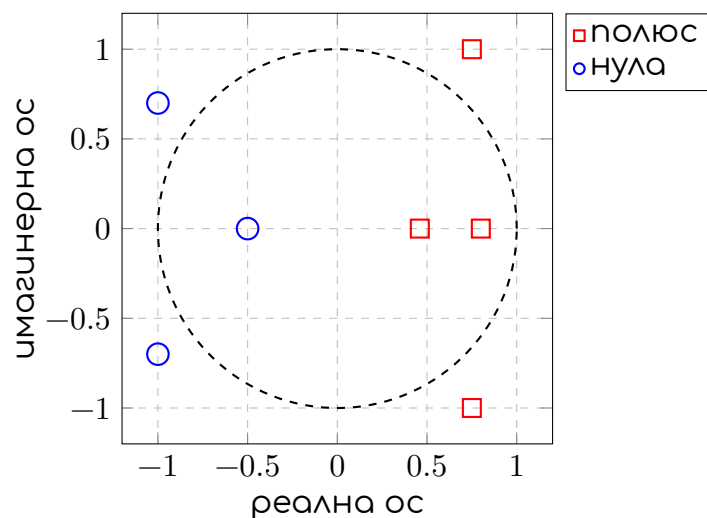
Това представяне е удобно, защото ни позволява да разбием \mathcal{H} на произведение от по-прости предавателни функции:

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

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

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

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



Фигура A.1.1: Полюс-нула графика

Фигура A.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.

A.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.

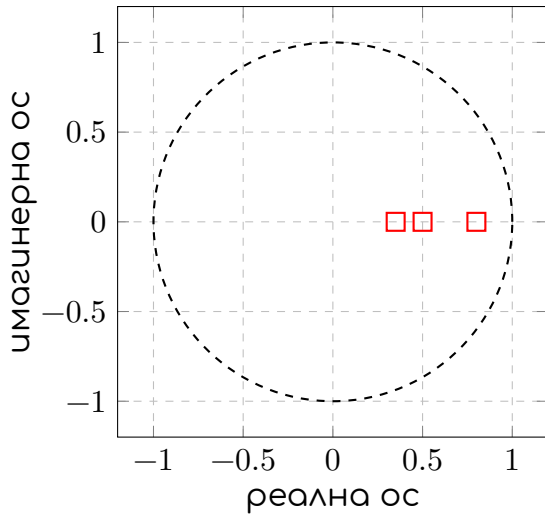
Нека имаме IIR филтър от първи ред:

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

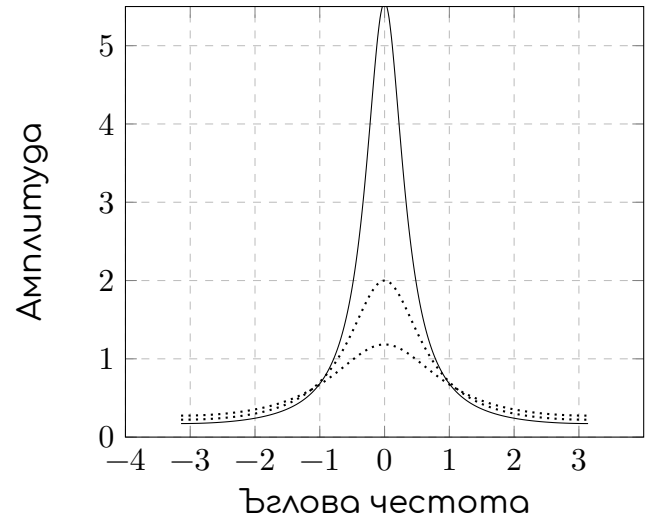
$$\mathcal{H} = G \frac{1}{1 - a_1 z^{-1}},$$

$$G = b_0$$

Тъй като имаме само един полюс, то следва, че a_1 е реално, 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.

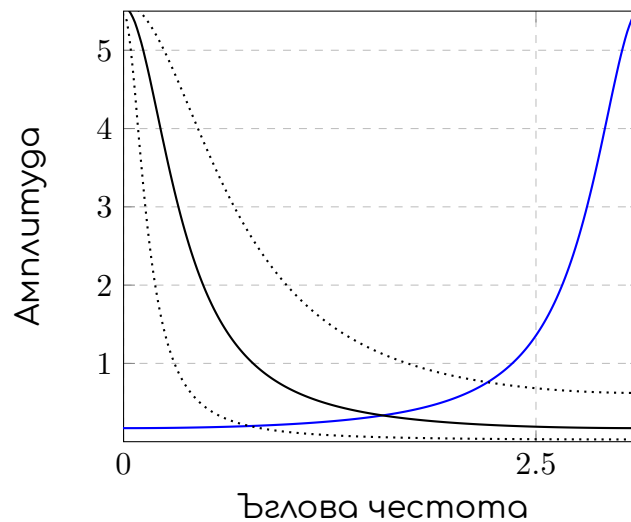


(а) Полус-нула графика за \mathcal{H}



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

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



Фигура A.2.2: Действие на IIR 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 .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 :

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

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

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

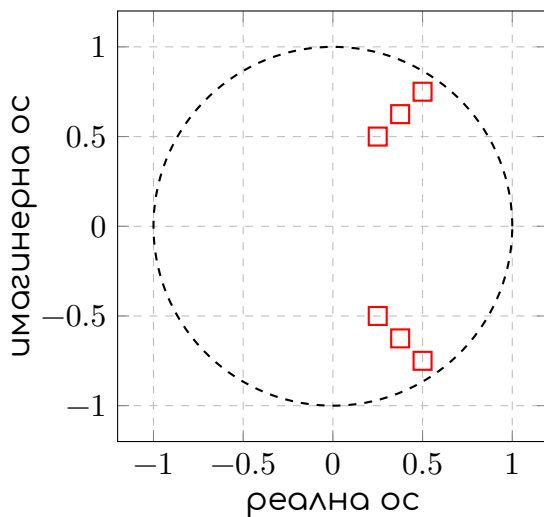
На Фигура .2.2 се вижда графиката на $g(\omega)$ за различни стойности на a_1 и b_0 . Такива филтри се наричат **резонатори**, тъй като честотите във върха на графиката ще се усилят. Резонаторите се описват главно чрез своята **амплитуда** - височината на максимума, **честота** - къде е върхът върху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

В случая на IIR 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. За да се преместим нанякъде по честотната ос, 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.

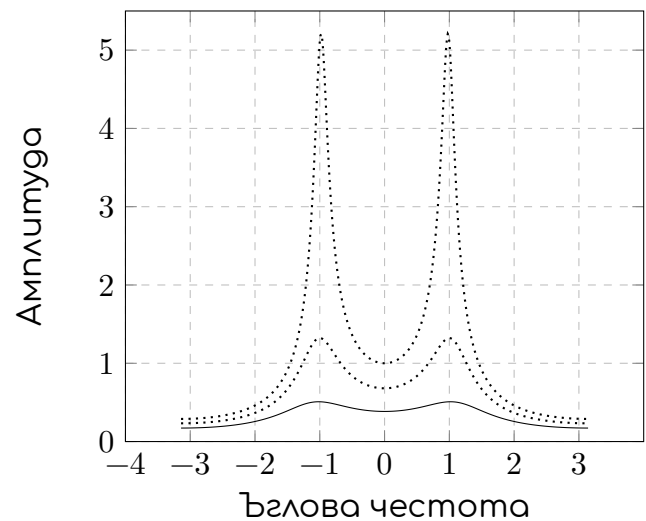
$$y[n] = b_0 x[n] + a_1 y[n-1] + a_2 y[n-2]$$

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

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

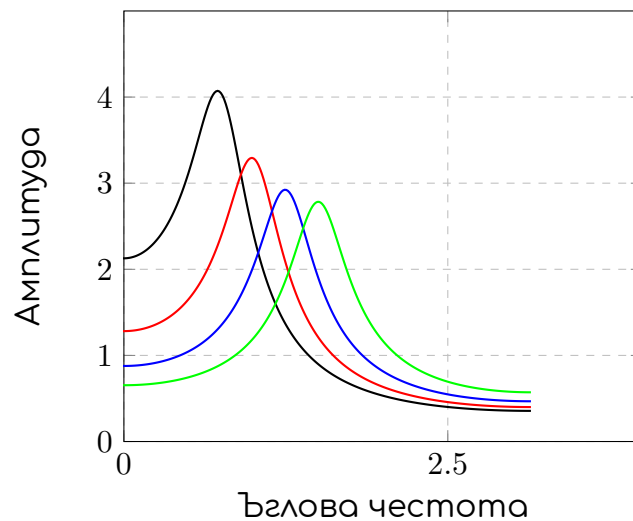


(а) Полус-нула графика за \mathcal{H}



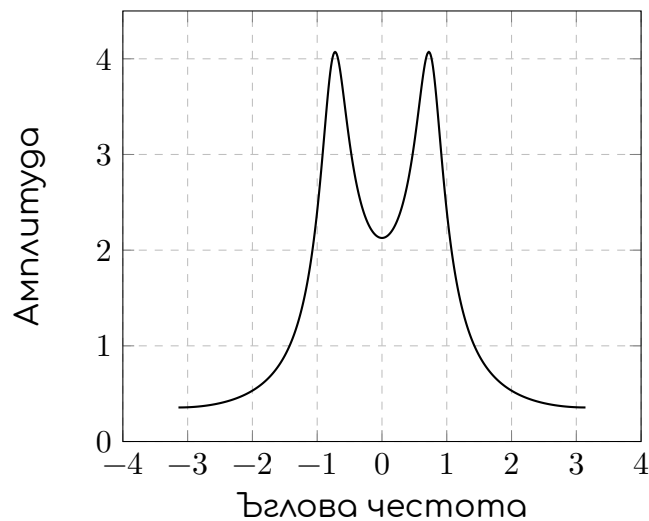
(б) Графика на $g(\omega)$ в $-\pi, \pi$

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

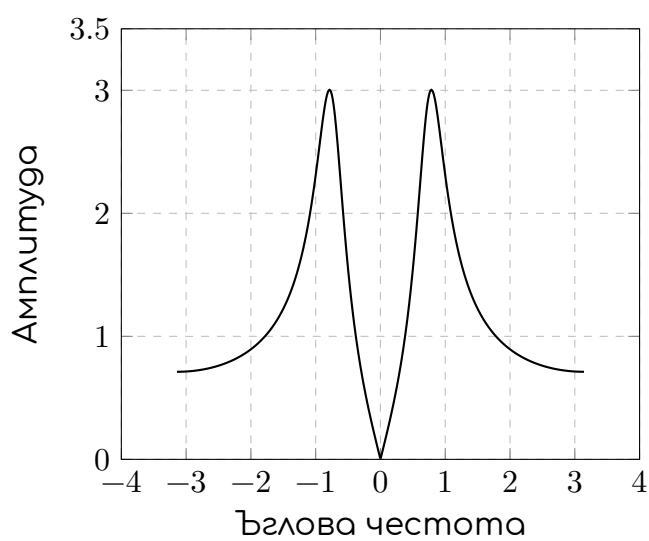


Фигура А.2.4: Графика на g с отдалечаващи се от реалната ос полюси

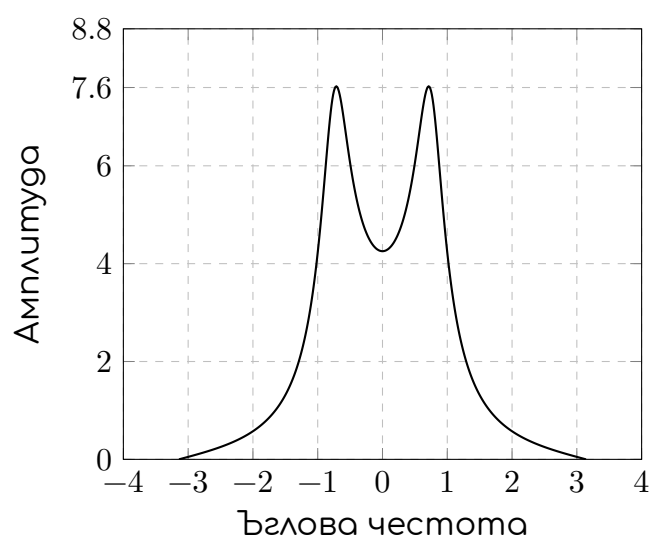
Местенето на полюсите по-далеч от реалната ос, раздалечава върховете по честотната лента, както се вижда на Фигура А.2.4



(a) $b_0 = 1, b_1 = 0$



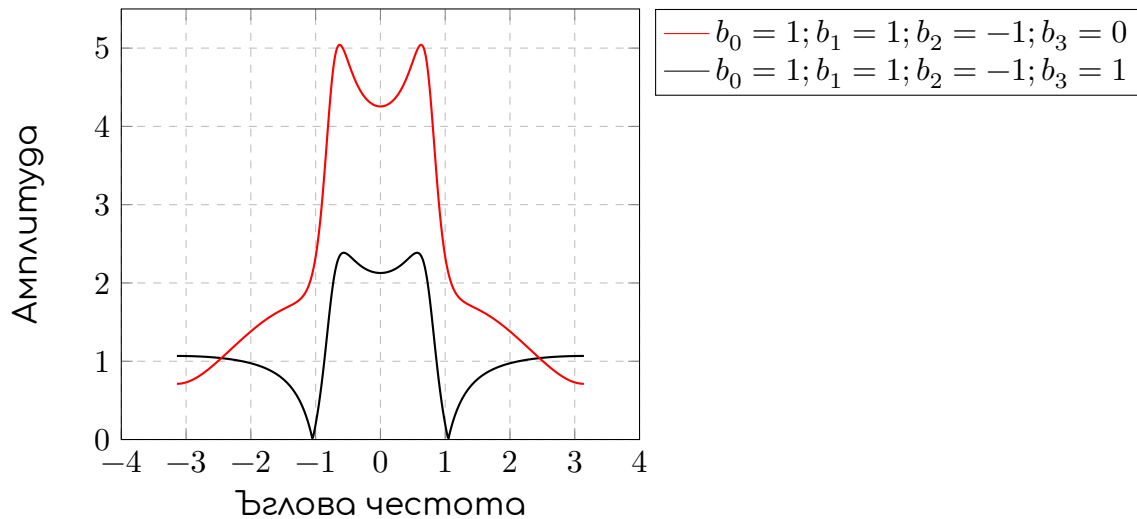
(б) $b_0 = 1, b_1 = 1$



(в) $b_0 = 1, b_1 = -1$

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

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



Фигура А.2.6: Действие на IIR 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 А.2.6. Тези нули се наричат **антирезонанси**.

Използвайки вида на IIR филтрите от първи и втори ред, можем да разложим даден сложен филтър \mathcal{H} на произведение на 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 и свойствата на Фурие преобразуванията ни дават вида и във времевия домейн.

Приложение Б

Приложение към Емоции в реч

Пример 1. $\mathcal{V}(z)$ за $N = 2$

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

$$= z \left[\frac{2}{1+r_G}, -\frac{2r_G}{1+r_G} \right] \hat{Q}_1 \hat{Q}_2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} =$$

$$= z \left[\frac{2}{1+r_G}, -\frac{2r_G}{1+r_G} \right] \begin{bmatrix} \frac{1}{1+r_1} & \frac{-r_1}{1+r_1} \\ \frac{-r_1 z^{-1}}{1+r_1} & \frac{z^{-1}}{1+r_1} \end{bmatrix} \begin{bmatrix} \frac{1}{1+r_2} & \frac{-r_2}{1+r_2} \\ \frac{-r_2 z^{-1}}{1+r_2} & \frac{z^{-1}}{1+r_2} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} =$$

$$= 2z \left[\frac{1+r_G r_1 z^{-1}}{(1+r_G)(1+r_1)}, -\frac{r_1+r_G z^{-1}}{(1+r_G)(1+r_1)} \right] \begin{bmatrix} \frac{1}{1+r_2} & \frac{-r_2}{1+r_2} \\ \frac{-r_2 z^{-1}}{1+r_2} & \frac{z^{-1}}{1+r_2} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} =$$

$$= 2z \left[\frac{1+r_G r_1 z^{-1} + r_1 r_2 z^{-1} + r_G r_2 z^{-2}}{(1+r_G)(1+r_1)(1+r_2)}, -\frac{r_2 + r_G r_1 r_2 z^{-1} + r_1 z^{-1} + r_G z^{-2}}{(1+r_G)(1+r_1)(1+r_2)} \right] \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

\Leftrightarrow

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

□