

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	Загладено опростяване	5
2.2.1	Преминаване от една тръба в друга	7
2.2.2	Ограничения при устните	10
2.2.3	Ограничения при глотиса	11
2.2.4	Общ вид на \mathcal{V}	11
2.3	Характеристики	11
2.3.1	Избор	11
2.3.2	Извличане	11
2.4	Класификация	11
2.5	Резултати	11
3	Грубо в мозъка	12
3.1	Характеристики	12
3.1.1	Избор	12
3.1.2	Извличане	12
3.2	Класификация	12
3.3	Резултати	12
4	Двойната звезда	13
4.1	Резултати	13
5	Големият портрет	14

Глава 1

Нулева зона

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



Глава 2

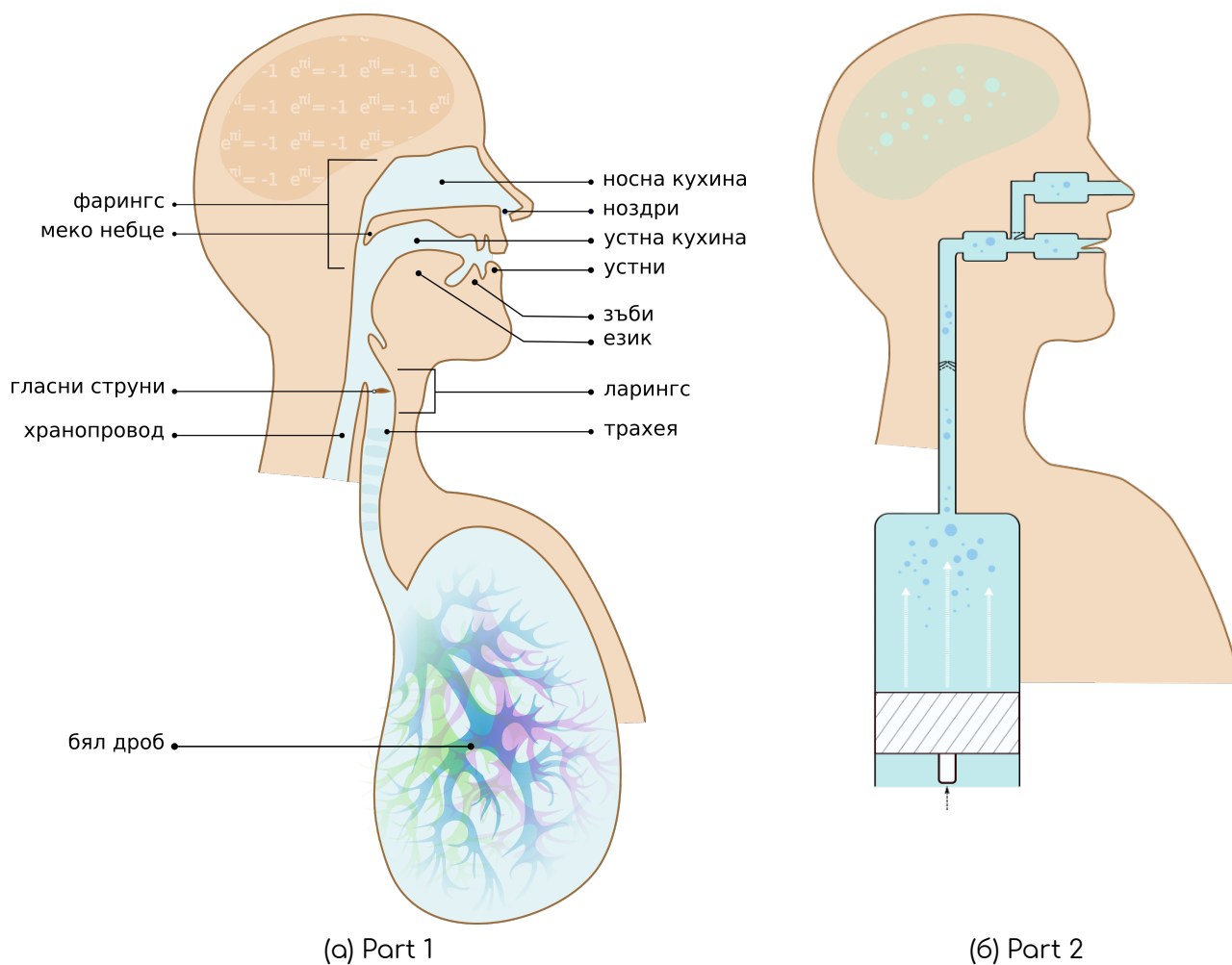
Емоции в реч

2.1 Физика на тъгата

Вокален тракт е общото название на кухините над ларингса (гръкляна), през които минава въздуха при генериране на реч. При хората той се състои от ларингеална кухина, фарингс, 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] и 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 Фиг.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 междуребрените 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 повта-



Фигура 2.1.1: Гррррр

ря 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 презгради, като устни и зъби, и в крайна сметка напуска системата през някой от отворите.

При целият 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 начинът, 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 презграда, което се освобождава рязко.

Ако разгледаме вокалния тракт и носната кухина като свързани тръби с непостоянно 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 онава в тетрадката. Това много прилича на свирене на духов инструмент. Честотите на които се получава резонанс зависи 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 Загладено опростяване

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

В случая на гласната "ъ", глотиса 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 . Сигналят, 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 $\mathcal{Y}(z) = \mathcal{G}(z) \mathcal{V}(z) \mathcal{R}(z)$, където

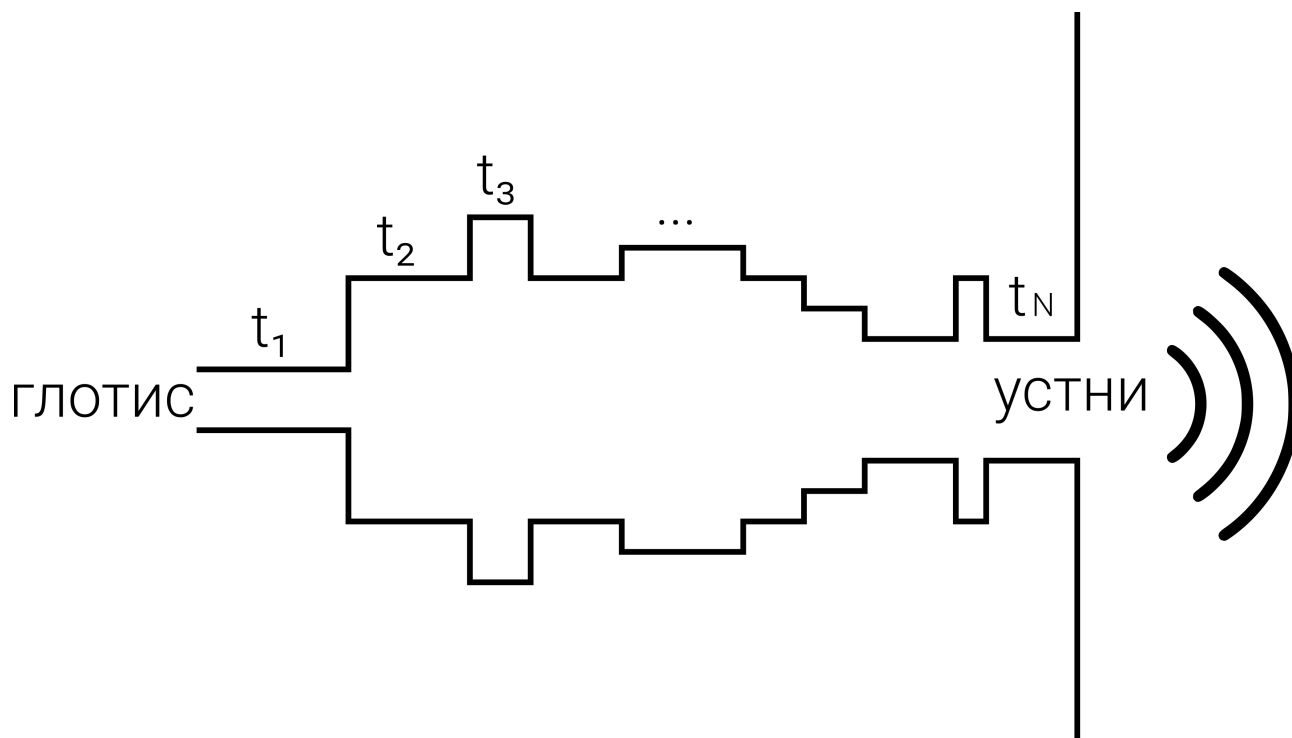
$$\mathcal{G}(z) = \mathcal{Z}\{g(t)\}$$

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

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

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

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



Фигура 2.2.1: Приближение на вокалния тракт с N тръби

Нека устната кухина е приближена с N на брой тръби с постоянно напречно сечение, както е показано на (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
5. p - налягане - можем да си мислим за него като "колко сила е нужна, за да се преместят частитиците

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 устните, а с u_G обемната скорост съответстваща първата точка на t_1 , тоест $u_G(t) = u_1(0, t)$. Тогава можем да опишем отношението на u_L и u_G по следния начин: обемната скорост при устните е резултат от прилагането на филтъра на вокалния тракт върху сигнала от глотиса $u_L(t) = u_G(t) * v(t) \longleftrightarrow v(t) = \frac{u_L}{u_G}$. Или изразено в честотния домейн като:

$$\mathcal{V} = \frac{u_L}{u_G}$$

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 (2.2.3a) в (2.2.2a), 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

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

\Leftrightarrow

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

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 **коэффициент 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 разглеждаме 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 (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) и има вида:

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

а L_r и R_r са константи, определени от размера на отвора на устата. При ниски честоти $Z_L(z) \approx 0$, което значи, че съпротивлението на устните е почти нулево. За реални приложения стойностите на L_r и R_r са така избрани, че 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.

Ако приемем, че Z_L е реално число, можем да разгледаме уравнението (2.2.13) във времевия домейн, ще получим:

$$\begin{aligned} p_L(t) &= Z_L u_L(t) \\ \frac{\rho c}{A_L} [u_N^+(t - \tau_N) + u_N^-(t + \tau_N)] &= Z_L [u_N^+(t - \tau_N) - u_N^-(t + \tau_N)] \\ u_N^-(t + \tau_N) \frac{(\rho c + A_L Z_L)}{A_L} &= u_N^+(t - \tau_N) \frac{(A_L Z_L - \rho c)}{A_L} \\ u_N^-(t + \tau_N) &= -r_L u_N^+(t - \tau_N), \text{ където} \\ r_L &= \left(\frac{\rho c - A_L Z_L}{\rho c + A_L Z_L} \right) \end{aligned} \quad (2.2.14)$$

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 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 и в такъв случай Z_L трябва да се замени с диференциално уравнение, описващо връзката между p_N и u_N .

2.2.3 Ограничения при глотиса

2.2.4 Общ вид на \mathcal{V}

2.3 Характеристики

2.3.1 Избор

2.3.2 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, нахъсваме масива с данните на отделни застъпващи се фреймове, в рамките на които сигналът 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 глотиса. 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 смплите 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]

2.4 Класификация

2.5 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