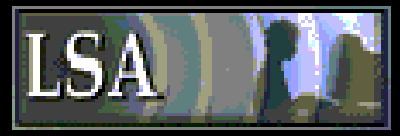
Statistical analysis of acoustical parameters in the voice of children with juvenile dysphonia

Miklós Gábriel Tulics, Ferenc Kazinczi and Klára Vicsi

BUTE

Laboratory of Speech Acoustics - Department of Telecommunications and Media Informatics



Goal

Goal

To build an automatic tool that can recognize disordered speech, which is generally caused by the defect of the vocal organs.

Goal

To build an automatic tool that can recognize disordered speech, which is generally caused by the defect of the vocal organs.

This is done by classification methods separating healthy from disordered speech. In this case the training data should statistically cover the data we want to recognize (testing data).

Goal

To build an automatic tool that can recognize disordered speech, which is generally caused by the defect of the vocal organs.

This is done by classification methods separating healthy from disordered speech. In this case the training data should statistically cover the data we want to recognize (testing data).

Most of the research in this field focuses mainly on adult voices, but what about the children's voices?

Goal

To build an automatic tool that can recognize disordered speech, which is generally caused by the defect of the vocal organs.

This is done by classification methods separating healthy from disordered speech. In this case the training data should statistically cover the data we want to recognize (testing data).

Most of the research in this field focuses mainly on adult voices, but what about the children's voices?

- **Main question: Is it necessary to build a completely different system in order to automatically recognize functional dysphonia (FD) in children's cases or is it possible to train the system with healthy and pathological voices of adults?**

What is functional dysphonia?

What is functional dysphonia?

- **Dysphonia is the disorder of the articulation as a complex function. It is a pathological condition showing varied based symptoms due to several etiologic factors and pathogenesis diversity.**
- **Functional dysphonia (FD) is a multifaceted voice disorder. It refers to a voice problem in the absence of a physical condition.**
- **Juvenile dysphonia is when functional dysphonia occurs at an early age.**

What is functional dysphonia?

- **Dysphonia is the disorder of the articulation as a complex function. It is a pathological condition showing varied based symptoms due to several etiologic factors and pathogenesis diversity.**
- **Functional dysphonia (FD) is a multifaceted voice disorder. It refers to a voice problem in the absence of a physical condition.**
- **Juvenile dysphonia is when functional dysphonia occurs at an early age.**
- **Discomfort associated with dysphonia are: pressure on the neck, forced coughing, shortness of breath.**

What is functional dysphonia?

- **Dysphonia is the disorder of the articulation as a complex function. It is a pathological condition showing varied based symptoms due to several etiologic factors and pathogenesis diversity.**
- **Functional dysphonia (FD) is a multifaceted voice disorder. It refers to a voice problem in the absence of a physical condition.**
- **Juvenile dysphonia is when functional dysphonia occurs at an early age.**
- **Discomfort associated with dysphonia are: pressure on the neck, forced coughing, shortness of breath.**
- **The frequency of dysphonia among the 3-10-year-old population can be put between 20-30%. The data therefore suggest that almost every fourth or fifth child produces a pathological voice. The studies agree that dysphonia is more often found among boys than girls, the ratio being 70-30%.**

In this study

In this study

- **Sustained voice or continuous speech?**
  - Acoustic parameters like Jitter, Shimmer, HNR (Harmonics-to-Noise Ratio) in the automatic classification of results from the healthy and pathological voices are improved in a big extent using continuous speech

- **In the present study the differences and the similarities between adult and children's voice was analyzed statistically using continuous speech.**

In this study

- **Sustained voice or continuous speech?**
  - Acoustic parameters like Jitter, Shimmer, HNR (Harmonics-to-Noise Ratio) in the automatic classification of results from the healthy and pathological voices are improved in a big extent using continuous speech

- **In the present study the differences and the similarities between adult and children's voice was analyzed statistically using continuous speech.**

- **The acoustic parameters were also compared with two sample T-tests in the case of children, between healthy and pathological group.**

In this study

- **Sustained voice or continuous speech?**
  - Acoustic parameters like Jitter, Shimmer, HNR (Harmonics-to-Noise Ratio) in the automatic classification of results from the healthy and pathological voices are improved in a big extent using continuous speech

- **In the present study the differences and the similarities between adult and children's voice was analyzed statistically using continuous speech.**

- **The acoustic parameters were also compared with two sample T-tests in the case of children, between healthy and pathological group.**

- **Different approaches were carried out: acoustic parameters from (the Hungarian SAMPA) vowels [E] and [o] were extracted from adult and children's speech samples and compared by statistical analyses.**

- **Differences and similarities of healthy voice samples between the adult and child group was examined. At the beginning of our research male and female samples were treated together, seeing the difference we arrived at the conclusion that it is better to treat them separately.**

Participants and methods / Healthy Adults Speech Database

Participants and methods / Healthy Adults Speech Database

- **In this work a total of 84 recordings were used from healthy people (42 female and 42 male)**

Participants and methods / Healthy Adults Speech Database

- In this work a total of 84 recordings were used from healthy people (42 female and 42 male)
- Near field microphone (Monacor ECM-100), with Creative Soundblaster Audigy 2 NX: an outer USB sound card with 44,100Hz sampling rate, at a 16-bit linear coding.

Participants and methods / Healthy Adults Speech Database

- In this work a total of 84 recordings were used from healthy people (42 female and 42 male)
- Near field microphone (Monacor ECM-100), with Creative Soundblaster Audigy 2 NX: an outer USB sound card with 44,100Hz sampling rate, at a 16-bit linear coding.
- The duration of the recordings are about one minute each. Every patient had to read out aloud one of Aesop's Fables, “The North Wind and the Sun”. This folktale is frequently used in phoniatrics as an illustration of spoken language. It has been translated into several languages, Hungarian included.

Participants and methods / Juvenile dysphonia and Healthy Child Speech Database

FD child

Healthy child

Participants and methods / Juvenile dysphonia and Healthy Child Speech Database

- **20 healthy and 12 (1 female and 11 male) recordings from children diagnosed with juvenile dysphonia (furthermore referred as FD children).**

FD child

Healthy child

Participants and methods / Juvenile dysphonia and Healthy Child Speech Database

- **20 healthy and 12 (1 female and 11 male) recordings from children diagnosed with juvenile dysphonia (furthermore referred as FD children).**
- **Conditions: microphone (Monacor ECM-100), Creative Soundblaster Audigy 2 NX outer USB sound card, with 44,100Hz sampling rate and 16-bit linear coding.**

FD child

Healthy child

Participants and methods / Juvenile dysphonia and Healthy Child Speech Database

- **20 healthy and 12 (1 female and 11 male) recordings from children diagnosed with juvenile dysphonia (furthermore referred as FD children).**
- **Conditions: microphone (Monacor ECM-100), Creative Soundblaster Audigy 2 NX outer USB sound card, with 44,100Hz sampling rate and 16-bit linear coding.**
- **The duration of the recordings are about 20 seconds each. The children told a two versed poem.**

FD child

Healthy child

Participants and methods / Juvenile dysphonia and Healthy Child Speech Database

- **20 healthy and 12 (1 female and 11 male) recordings from children diagnosed with juvenile dysphonia (furthermore referred as FD children).**
- **Conditions: microphone (Monacor ECM-100), Creative Soundblaster Audigy 2 NX outer USB sound card, with 44,100Hz sampling rate and 16-bit linear coding.**
- **The duration of the recordings are about 20 seconds each. The children told a two versed poem.**
- **All the children from the database are aged 5 to 6. All the recordings were made in the presence of the children's parents.**

FD child

Healthy child

Participants and methods / Juvenile dysphonia and Healthy Child Speech Database

- **20 healthy and 12 (1 female and 11 male) recordings from children diagnosed with juvenile dysphonia (furthermore referred as FD children).**
- **Conditions: microphone (Monacor ECM-100), Creative Soundblaster Audigy 2 NX outer USB sound card, with 44,100Hz sampling rate and 16-bit linear coding.**
- **The duration of the recordings are about 20 seconds each. The children told a two versed poem.**
- **All the children from the database are aged 5 to 6. All the recordings were made in the presence of the children's parents.**
- **Both databases were annotated and segmented on phoneme level, using the SAMPA phonetic alphabet.**

FD child

Healthy child

Pre-processing methods

Pre-processing methods

- **Among the 14 Hungarian vowels, [E] and [o] are usually analysed in case of adults. There are approximately 50 [E] vowels in the tale that was read.**

- **In the case of the children, the vowel [o] is the poem's most frequent one, with 16 pieces, and there are only 9 pieces of the vowel [E]. The statistical analyses were made extracting the vowels [E], [o] from each database.**

Pre-processing methods

- **Among the 14 Hungarian vowels, [E] and [o] are usually analysed in case of adults. There are approximately 50 [E] vowels in the tale that was read.**
- **In the case of the children, the vowel [o] is the poem's most frequent one, with 16 pieces, and there are only 9 pieces of the vowel [E]. The statistical analyses were made extracting the vowels [E], [o] from each database.**

Parameters	Short description
Jitter_ddp	the average absolute difference between consecutive differences between consecutive periods, divided by the average period
Shimmer_ddp	the average absolute difference between consecutive differences between the amplitudes of consecutive periods
HNR	Harmonics-to-Noise Ratio quantifies noise in the speech signal, caused mainly due to incomplete vocal fold closure
MFCC1	the first component of the mel-frequency cepstral coefficients

Pre-processing methods

- **Among the 14 Hungarian vowels, [E] and [o] are usually analysed in case of adults. There are approximately 50 [E] vowels in the tale that was read.**
- **In the case of the children, the vowel [o] is the poem's most frequent one, with 16 pieces, and there are only 9 pieces of the vowel [E]. The statistical analyses were made extracting the vowels [E], [o] from each database.**
- **For the extraction of the acoustic parameters Praat software was used. The acoustic parameters were measured in the middle of the vowels.**

Parameters	Short description
Jitter_ddp	the average absolute difference between consecutive differences between consecutive periods, divided by the average period
Shimmer_ddp	the average absolute difference between consecutive differences between the amplitudes of consecutive periods
HNR	Harmonics-to-Noise Ratio quantifies noise in the speech signal, caused mainly due to incomplete vocal fold closure
MFCC1	the first component of the mel-frequency cepstral coefficients

Acoustic parameters

Acoustic parameters

Acoustic Parameters	Short description	Equation
Jitter_ddp	the average absolute difference between consecutive differences between consecutive periods, divided by the average period	$Jitter(ddp) = \frac{\frac{1}{N-2}\sum_{i=2}^{N-1}\lvert(T_{i+1}-T_i)-(T_i-T_{i-1})\rvert}{\frac{1}{N}\sum_{i=1}^{N}T_i}$ [%], Ti is the duration of the i-th interval and N is the number of intervals
Shimmer_ddp	the average absolute difference between consecutive differences between the amplitudes of consecutive periods	$Shimmer(ddp) = \frac{\frac{1}{N-2}\sum_{i=2}^{N-1}\lvert(A_{i+1}-A_i)-(A_i-A_{i-1})\rvert}{\frac{1}{N}\sum_{i=1}^{N}A_i}$ [%]
HNR	Harmonics-to-Noise Ratio quantifies noise in the speech signal, caused mainly due to incomplete vocal fold closure	$HNR = 10 \cdot lg\ \frac{E_H}{E_N}\ [dB]$
MFCC1	the first component of the mel-frequency cepstral coefficients	$c_{k-1} = \sum_{j=1}^{N} P_j \cos\left(\frac{\pi(k-1)(j-0,5)}{N}\right)$, N represents the number of spectral values and Pj the power in dB of the jth spectral value (k runs from 1 to N)

Statistical analyses

- **SPSS20.0 software was used**
- **Two sample T-tests were used for statistical significance testing.**
- **Where F tests showed significant variances of an acoustic parameter within the groups (with significance level 95% (α = 0.05), Welch's T-test was used.**
- **Our assumption is that the distributions are normal, but T tests are relatively robust to moderate violations of the normality assumption.**
- **The null hypothesis is that the means are equal.**

Results / Comparison of healthy and FD children

Results / Comparison of healthy and FD children

	[E]	[o]
	p - values	
Jitter_ddp	**0.018****	**0.000*****
Shimmer_ddp	**0.000*****	**0.000*****
mean_HNR	**0.072***	**0.003*****
MFCC1	**0.000*****	**0.000*****

Results / Comparison of healthy and FD children

	[E]	[o]
	p - values	
Jitter_ddp	**0.018****	**0.000*****
Shimmer_ddp	**0.000*****	**0.000*****
mean_HNR	**0.072***	**0.003*****
MFCC1	**0.000*****	**0.000*****

- * p < 0.1
- ** p < 0.05 -> indicates strong evidence against the null hypothesis, so the null hypothesis is rejected
- *** p < 0.01

Results / Comparison of healthy and FD child

The poem contains 9 instances of [E] and 16 instances of [o] sounds.

	[E]	[o]
	p - values	
Jitter_ddp	**0.018****	**0.000*****
Shimmer_ddp	**0.000*****	**0.000*****
mean_HNR	**0.072***	**0.003*****
MFCC1	**0.000*****	**0.000*****

- * p < 0.1
- ** p < 0.05 -> indicates strong evidence against the null hypothesis, so the null hypothesis is rejected
- *** p < 0.01

Results / Comparison of healthy adults and children /1

Results / Comparison of healthy adults and children /1

	Vowel									
	[o]					[E]				
	Child		Male			Child		Male		
	Mean	Std. Dev.	Mean	Std. Dev.	p-value	Mean	Std. Dev.	Mean	Std. Dev.	p-value
Jitter_ddp	1.095	0.740	1.448	1.533	**0.020****	1.414	1.084	1.986	1.791	**0.000*****
Shimmer_ddp	7.514	3.698	8.654	6.962	0.109	9.669	5.268	12.125	10.070	**0.003*****
mean_HNR	17.982	4.232	12.872	4.776	**0.000*****	13.262	3.914	8.337	4.068	**0.000*****
MFCC1	245.977	45.554	265.013	54.779	**0.000*****	175.224	32.372	208.357	51.887	**0.000*****

Results / Comparison of healthy adults and children /1

	Vowel									
	[o]					[E]				
	Child		Male			Child		Male		
	Mean	Std. Dev.	Mean	Std. Dev.	p-value	Mean	Std. Dev.	Mean	Std. Dev.	p-value
Jitter_ddp	1.095	0.740	1.448	1.533	**0.020****	1.414	1.084	1.986	1.791	**0.000*****
Shimmer_ddp	7.514	3.698	8.654	6.962	0.109	9.669	5.268	12.125	10.070	**0.003*****
mean_HNR	17.982	4.232	12.872	4.776	**0.000*****	13.262	3.914	8.337	4.068	**0.000*****
MFCC1	245.977	45.554	265.013	54.779	**0.000*****	175.224	32.372	208.357	51.887	**0.000*****

- * $p < 0.1$
- ** $p < 0.05$ -> indicates strong evidence against the null hypothesis, so the null hypothesis is rejected
- *** $p < 0.01$

Results / Comparison of healthy adults and children /2

Results / Comparison of healthy adults and children /2

	Vowel									
	[o]					[E]				
	Child		Female			Child		Female		
	Mean	Std. Dev.	Mean	Std. Dev.	p-value	Mean	Std. Dev.	Mean	Std. Dev.	p-value
Jitter_ddp	1.095	0.740	0.904	1.026	**0.013****	1.414	1.084	1.173	1.155	**0.011****
Shimmer_ddp	7.514	3.698	5.813	4.653	**0.000*****	9.669	5.268	7.755	5.717	**0.000*****
mean_HNR	17.982	4.232	17.109	5.223	0.370	13.262	3.914	12.764	4.408	0.128
MFCC1	245.977	45.554	249.780	47.506	0.347	175.224	32.372	177.965	43.429	0.320

Results / Comparison of healthy adults and children /2

	Vowel									
	[o]					[E]				
	Child		Female			Child		Female		
	Mean	Std. Dev.	Mean	Std. Dev.	p-value	Mean	Std. Dev.	Mean	Std. Dev.	p-value
Jitter_ddp	1.095	0.740	0.904	1.026	**0.013****	1.414	1.084	1.173	1.155	**0.011****
Shimmer_ddp	7.514	3.698	5.813	4.653	**0.000*****	9.669	5.268	7.755	5.717	**0.000*****
mean_HNR	17.982	4.232	17.109	5.223	0.370	13.262	3.914	12.764	4.408	0.128
MFCC1	245.977	45.554	249.780	47.506	0.347	175.224	32.372	177.965	43.429	0.320

- * $p < 0.1$
- ** $p < 0.05$ -> indicates strong evidence against the null hypothesis, so the null hypothesis is rejected
- *** $p < 0.01$

Results / Comparison of healthy adults and

Differences exist in the examined acoustical parameters even between healthy child and healthy adult groups.

A decision system that inquiries child's voice trained with adult voice samples would likely detract erroneous conclusions

	Vowel									
	[o]					[ɛ]				
	Child		Female			Child		Female		
	Mean	Std. Dev.	Mean	Std. Dev.	p-value	Mean	Std. Dev.	Mean	Std. Dev.	p-value
Jitter_ddp	1.095	0.740	0.904	1.026	**0.013****	1.414	1.084	1.173	1.155	**0.011****
Shimmer_ddp	7.514	3.698	5.813	4.653	**0.000*****	9.669	5.268	7.755	5.717	**0.000*****
mean_HNR	17.982	4.232	17.109	5.223	0.370	13.262	3.914	12.764	4.408	0.128
MFCC1	245.977	45.554	249.780	47.506	0.347	175.224	32.372	177.965	43.429	0.320

- * $p < 0.1$
- ** $p < 0.05$ -> indicates strong evidence against the null hypothesis, so the null hypothesis is rejected
- *** $p < 0.01$

Results / Comparison of healthy males and females

Results / Comparison of healthy males and females

	Vowel									
	[o]					[E]				
	Female		Male			Female		Male		
	Mean	Std. Dev.	Mean	Std. Dev.	p-value	Mean	Std. Dev.	Mean	Std. Dev.	p-value
Jitter_ddp	0.904	1.026	1.448	1.533	**0.000*** **	1.173	1.155	1.986	1.791	**0.000*** **
Shimmer_ddp	5.813	4.653	8.654	6.962	**0.000*** **	7.755	5.717	12.125	10.070	**0.000*** **
mean_HNR	17.109	5.223	12.872	4.776	**0.000*** **	12.764	4.408	8.337	4.068	**0.000*** **
MFCC1	249.780	47.506	265.013	54.779	**0.007*** **	177.965	43.429	208.357	51.887	**0.000*** **

Results / Comparison of healthy males and females

	Vowel									
	[o]					[E]				
	Female		Male			Female		Male		
	Mean	Std. Dev.	Mean	Std. Dev.	p-value	Mean	Std. Dev.	Mean	Std. Dev.	p-value
Jitter_ddp	0.904	1.026	1.448	1.533	**0.000*** **	1.173	1.155	1.986	1.791	**0.000*** **
Shimmer_ddp	5.813	4.653	8.654	6.962	**0.000*** **	7.755	5.717	12.125	10.070	**0.000*** **
mean_HNR	17.109	5.223	12.872	4.776	**0.000*** **	12.764	4.408	8.337	4.068	**0.000*** **
MFCC1	249.780	47.506	265.013	54.779	**0.007*** **	177.965	43.429	208.357	51.887	**0.000*** **

- * p < 0.1
- ** p < 0.05 -> indicates strong evidence against the null hypothesis, so the null hypothesis is rejected
- *** p < 0.01

Results / Comparison of healthy males and fema

It is reasonable to separate male and female samples when we have small dataset.

	Vowel									
	[o]					[E]				
	Female		Male			Female		Male		
	Mean	Std. Dev.	Mean	Std. Dev.	p-value	Mean	Std. Dev.	Mean	Std. Dev.	p-value
Jitter_ddp	0.904	1.026	1.448	1.533	**0.000*** **	1.173	1.155	1.986	1.791	**0.000*** **
Shimmer_ddp	5.813	4.653	8.654	6.962	**0.000*** **	7.755	5.717	12.125	10.070	**0.000*** **
mean_HNR	17.109	5.223	12.872	4.776	**0.000*** **	12.764	4.408	8.337	4.068	**0.000*** **
MFCC1	249.780	47.506	265.013	54.779	**0.007*** **	177.965	43.429	208.357	51.887	**0.000*** **

- * p < 0.1
- ** p < 0.05 -> indicates strong evidence against the null hypothesis, so the null hypothesis is rejected
- *** p < 0.01

Conclusions

Conclusions

- **Variations of Jitter and Shimmer values with HNR and MFCC1 are good indicators to separate healthy and FD voices in case of children as well.**

Conclusions

- **Variations of Jitter and Shimmer values with HNR and MFCC1 are good indicators to separate healthy and FD voices in case of children as well.**

- **Healthy samples of children and adult voices were compared giving the clear conclusion that differences exist in the examined acoustical parameters even between healthy child and healthy adult groups.**

Conclusions

- Variations of Jitter and Shimmer values with HNR and MFCC1 are good indicators to separate healthy and FD voices in case of children as well.
- Healthy samples of children and adult voices were compared giving the clear conclusion that differences exist in the examined acoustical parameters even between healthy child and healthy adult groups.
- It is necessary to carry out the investigations separately on children's voices as well, we cannot use adult voices to make any conclusions to children's voices.

Conclusions

- Variations of Jitter and Shimmer values with HNR and MFCC1 are good indicators to separate healthy and FD voices in case of children as well.
- Healthy samples of children and adult voices were compared giving the clear conclusion that differences exist in the examined acoustical parameters even between healthy child and healthy adult groups.
- It is necessary to carry out the investigations separately on children's voices as well, we cannot use adult voices to make any conclusions to children's voices.
- In order to build an automatic decision making system that recognizes FD it is advisable to train the system separately for adult males, adult females and children.

Thank you for your attention!

tulics@tmit.bme.hu