AKADEMIA GÓRNICZO-HUTNICZA
IM. STANISŁAWA STASZICA W KRAKOWIE

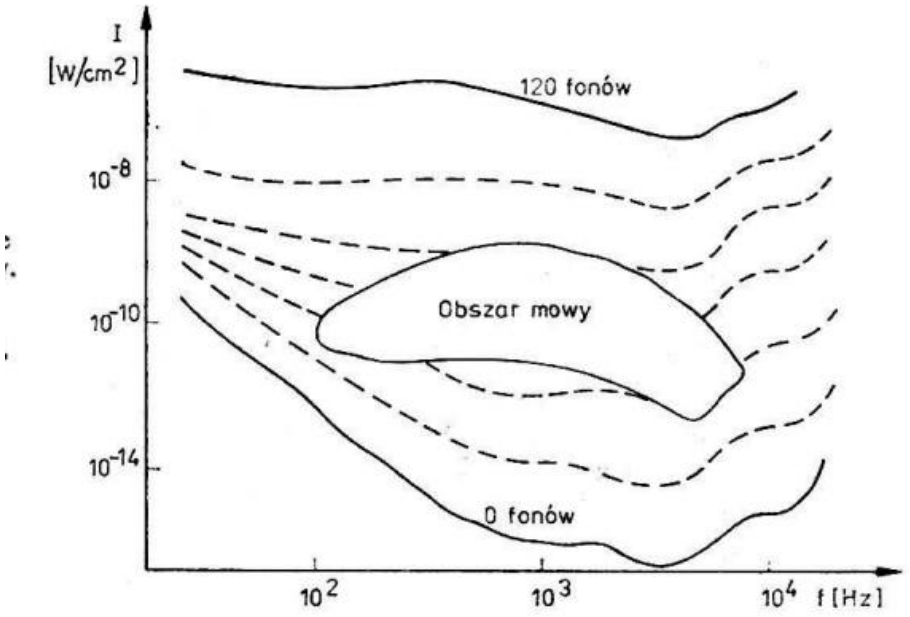
Odbiór mowy naturalnej i syntetycznej przez człowieka.

Agnieszka Koszany
Jagna Chronowska

Wydział Inżynierii Mechanicznej i Robotyki

Kraków, 11 stycznia 2016 r.

Słuch a mowa

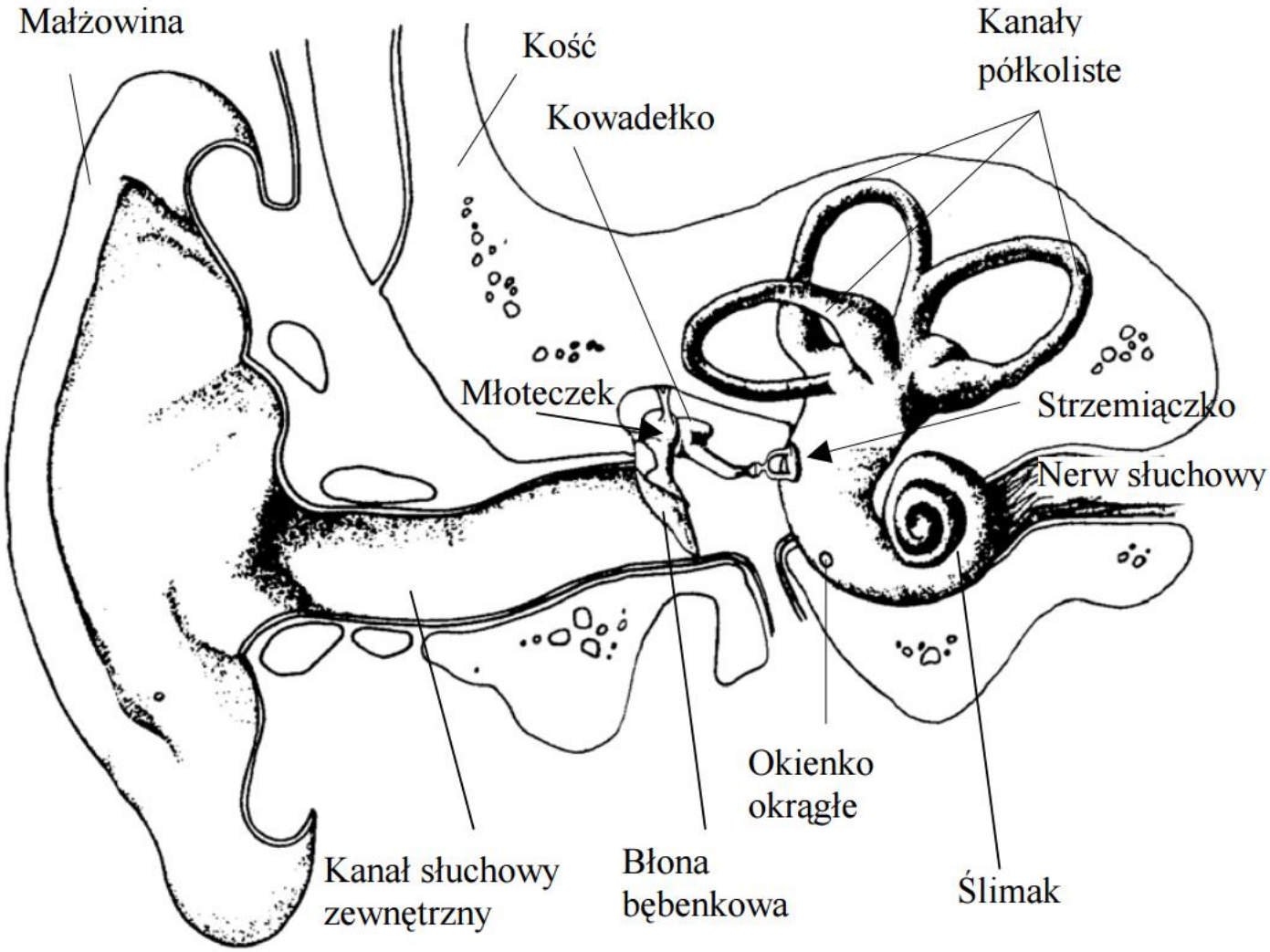


W procesie ewolucji mowa dostosowała się do percepcji narządu słuchu.

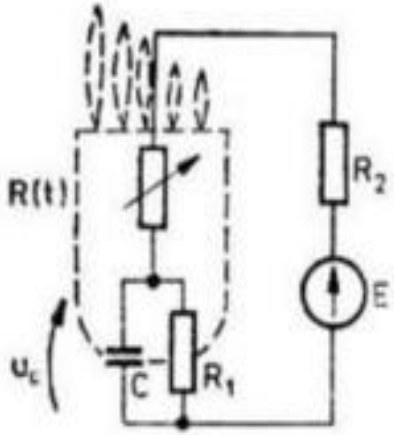
Poziomy percepcji mowy

- Aerodynamiczna – drgania powietrza w przewodzie słuchowym,
- Akustyczno-mechaniczna – przenoszenie drgań od błony bębenkowej do ślimaka,
- Neurologiczna – przenoszenia i przetwarzanie impulsów w ośrodkowym układzie nerwowym,
- Psychologiczna – rozpoznanie i zrozumienie przekazanej informacji.

Budowa ucha


Małżowina
Kość
Kowadełko
Kanały
półkoliste
Młoteczek
Strzemiączko
Nerw słuchowy
Okienko
okrągłe
Kanał słuchowy
zewnętrzny
Błona
bębenkowa
Ślimak

Schemat elektryczny komórki rzęskowatej



R(x) – zmienna rezystancja ciałka Hansena

R_1 – oporność środowiska

R_2 – oporność błony bębenkowej

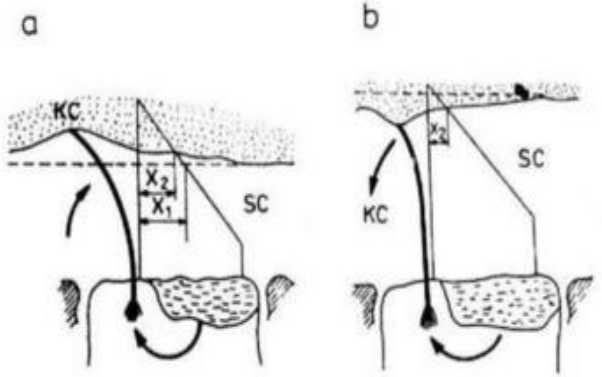
C - pojemność błony bębenkowej

E – biologiczne źródło zasilania

Wewnątrzkomórkowe sprzężenie zwrotne

Dla dźwięku o dużej mocy kinocilium odpycha błonę nakrywkową, co powoduje zmniejszenie wrażliwości słuchu.

Dla cichych dźwięków błona zostaje przyciągnięta, więc zwiększa się obszar ugięcia. Słuch staje się bardziej czuły.



Specyfika percepcji mowy

- Dźwięki mowy są lepiej dekodowane przez lewą półkulę mózgu – lepsza percepcja prawego ucha
- Różna impedancja ośrodków
- Wpływ na percepcję ruchu twarzy i warg
- Interpretacja treści jest indywidualną kwestią

Social perception of male and female computer synthesized speech. University of Pittsburgh (2003)

	Human voice comparisons						Synthetic voice comparisons					
	Male human voice		Female human voice		Contrast		Male synthetic voice		Female synthetic voice		Contrast	
					t	*r*					*t*	*r*
	Male listener	Female listener	Male listener	Female listener			Male listener	Female listener	Male listener	Female listener		
Perceptions of speaker												
Knowledgeable	14.68	15.00	15.18	15.71	0.32	0.03	14.12	15.65	14.70	14.80	2.22**	0.22[b]
Truthful	9.58	10.8	10.55	11.12	1.47*	0.15	9.16	10.42	9.74	10.40	1.37*	0.14
Involved	13.08	13.88	14.32	14.96	0.24	0.02	13.08	14.86	13.43	13.40	2.78**	0.27[b]
Powerful	7.56	8.12	8.45	8.71	0.64	0.07	9.56	9.69	8.74	8.48	0.84	0.08
Accurate	11.68	12.08	11.59	12.04	0.13	0.01	11.52	12.15	10.91	11.76	0.56	0.06
Perceptions of message												
Captivating	7.61	9.16	8.95	9.50	0.98	0.10	8.60	10.92	8.52	8.76	2.06**	0.21[b]
Clear	11.73	12.16	12.04	12.04	1.18	0.12	11.56	12.32	11.09	12.24	1.08	0.11
Convincing	4.50	4.88	5.19	4.83	2.12**	0.21[b]	4.20	5.12	4.30	4.48	2.14**	0.21[b]
Simple	4.35	4.76	4.45	4.17	2.50**	0.25[b]	4.12	4.00	4.35	4.20	0.11	0.01
Effectiveness of message												
Good	6.27	6.76	6.71	6.96	0.90	0.09	6.20	7.00	6.00	6.48	1.22	0.12
Wise	6.31	6.64	6.81	6.79	1.26	0.13	6.24	6.77	6.30	6.60	0.84	0.08
Positive	6.50	6.92	7.33	7.25	1.73**	0.18	6.52	7.44	6.52	6.72	2.52**	0.25[b]
Harmful	3.54	2.42	2.86	2.71	3.59**	0.35[a]	3.08	2.73	3.00	3.40	2.80**	0.27[b]
Unconvincing	3.50	3.12	2.81	3.17	2.12**		3.80	2.88	3.70	3.52	2.14**	0.21[b]
Ineffective	3.69	3.24	2.67	3.04	1.87**	0.19	3.48	3.50	3.78	3.48	0.74	0.07
Perceptions of speech qualities												
Soft	4.35	4.08	3.73	3.54	0.33	0.03	3.04	3.27	3.65	3.72	0.68	0.07
Squeaky	3.54	3.28	4.27	3.75	1.17	0.12	3.40	3.46	4.65	4.64	0.32	0.03
Slow	4.19	4.44	4.09	4.00	1.27	0.13	3.76	3.73	2.87	3.16	1.21	0.12
Unaccented	5.50	5.36	5.32	5.54	1.19	0.12	4.08	4.31	3.43	4.00	1.13	0.11
Not long	3.54	3.48	3.18	3.54	1.38*	0.14	2.72	3.11	3.22	3.20	1.67**	0.17
Less nasal	4.12	4.44	4.95	4.96	0.94	0.10	3.48	3.42	2.83	3.28	1.56*	0.16
Lively	2.65	2.72	3.45	3.75	0.81	0.08	1.64	1.73	1.70	1.96	0.60	0.06

Power determinations are from Cohen's (1988) power tables.

[a] Denotes power level >0.60, <0.80.

[b] Denotes power level >0.25, <0.60.

* Denotes a marginally significant contrast at $P<0.10$.

** Denotes a significant contrast at $P<0.05$.

Wnioski z badań

Syntezowana mowa damska jest lepiej odbierana przez kobiety, analogicznie męska lepiej odbierana jest przez mężczyzn.

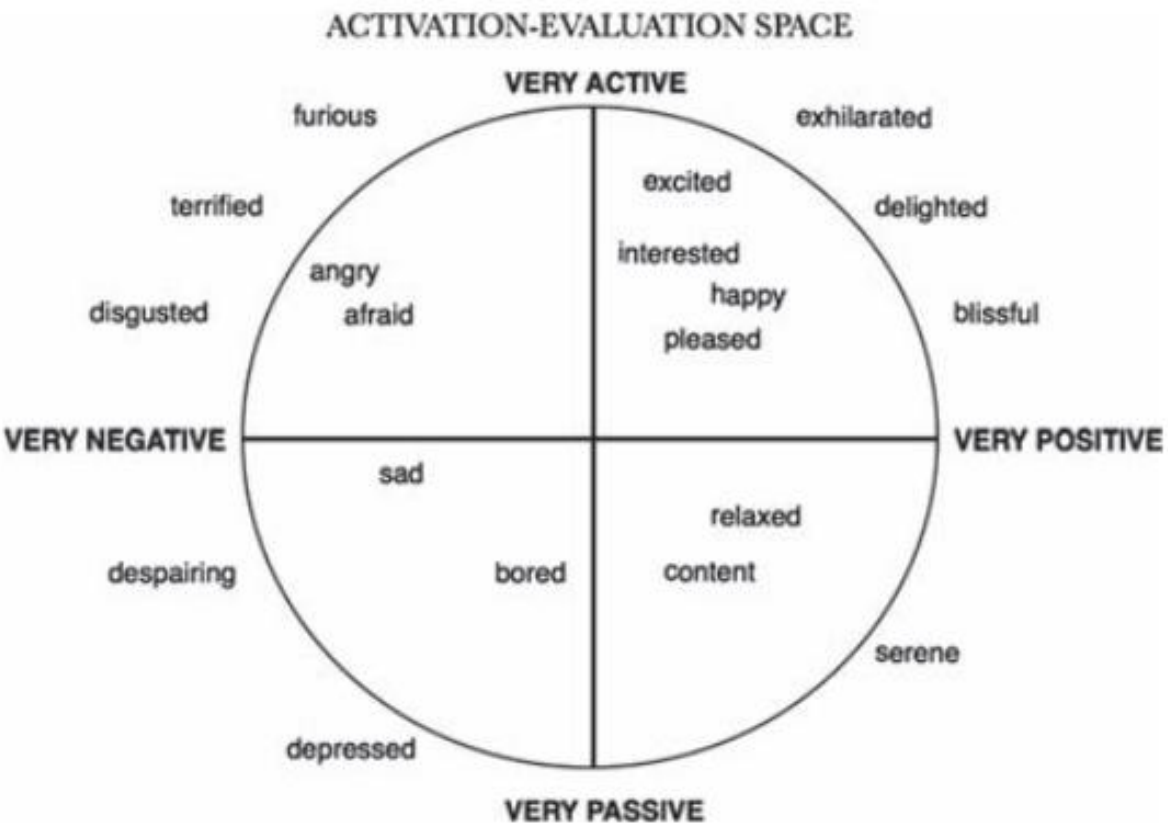
Oceny poszczególnych parametrów mowy naturalnej i syntetycznej nie były bardzo zróżnicowane- poprzez syntezę mowy można równie dobrze przekazywać informacje.

Exploring expressivity and emotion with artificial voice and speech technologies (2013):

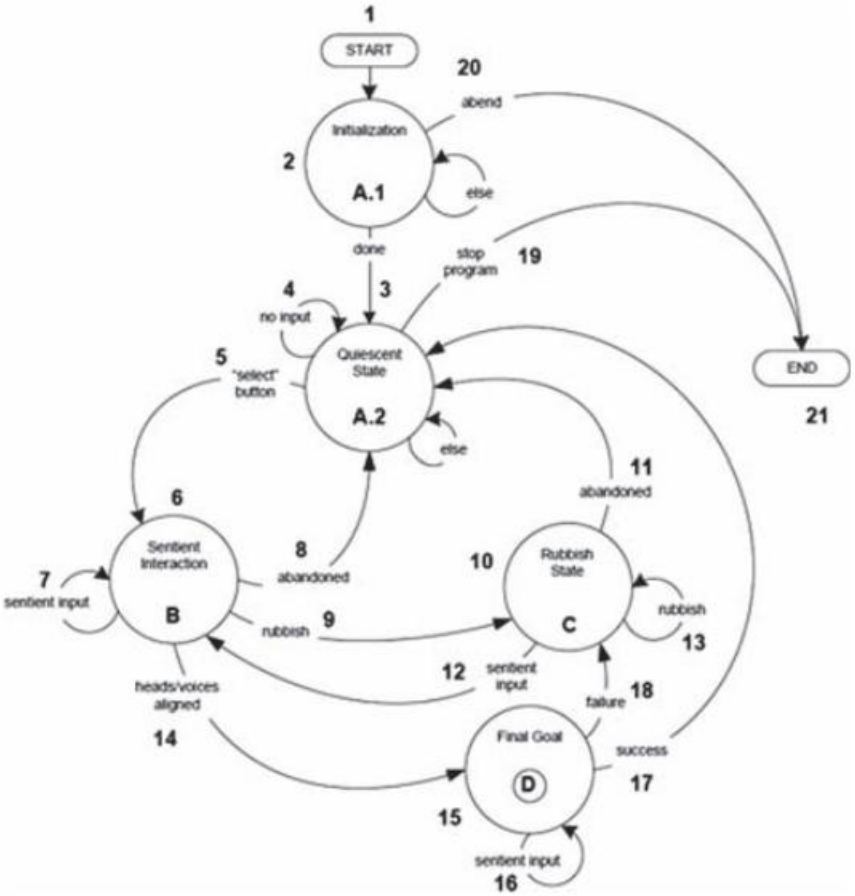
Badanie opierało się na rozmowie człowieka komputerem. Specjalny program stworzony do tego celu potrafił przeprowadzać inteligentną rozmowę, używając do tego specjalny algorytm.

Celem badania była obserwacja badanych osób w interakcji z komputerem, oraz określenie emocji oraz stosunku ludzi do syntezowanej mowy.

Skala emocji (każdą emocję można zbudować z 4 głównych)



Algorytm przeprowadzania rozmowy



Przebieg eksperymentu

Aby nadać poczucie rozmowy z czymś bardziej ludzkim, za „interfejs” służył kask motocyklowy oraz żarówka święcąca podczas wypowiadania fraz poprzez komputer.

Manipulacja frazami została także przedstawiona graficznie. Możliwa była manipulacja głośnością, długością wypowiadania wybranych słów. Mowa jednak miała zostać dalej „wypowiadana bez uczuć”.

Wygląd symulatora

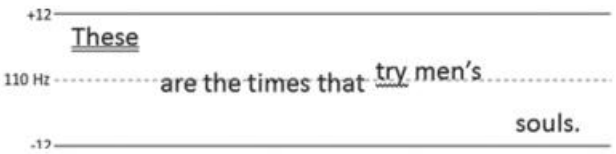


Figure 9. More elaborate prosodic specifier.

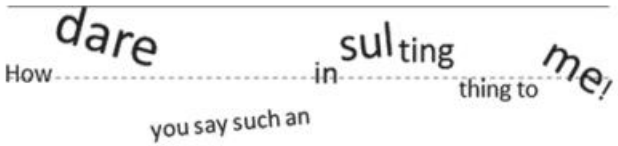
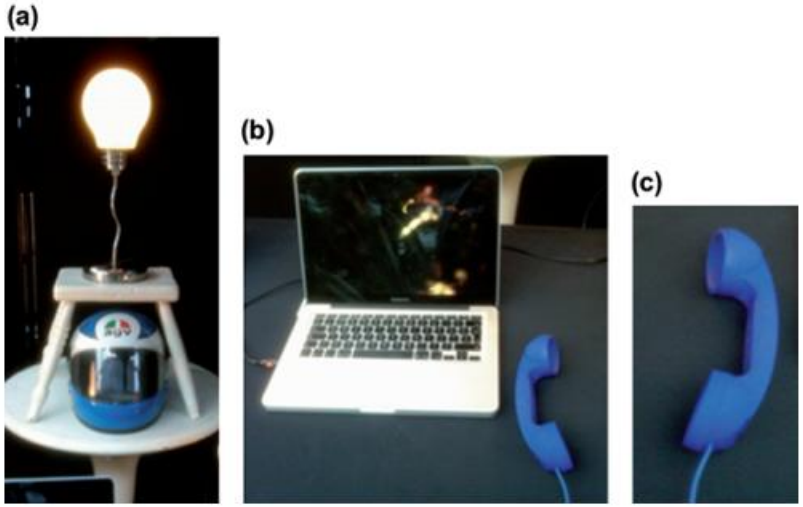


Figure 10. Graphical text manipulation.



Wnioski wynikające z badania

- Badane osoby podświadomie nadpisywały uczucia do syntetycznej mowy.
- Istotne w odbiorze była szybkość i głośność wypowiadanych zdań- badani odczuwali, że ich rozmówca prawdopodobnie jest znudzony bądź zły.
- Rozmówcom towarzyszyły głównie pozytywne emocje- percepcja syntetycznej mowy, rozmowa z komputerem traktowana jako pozytywne doświadczeni.

My voice and me, 2013

Oliver Curry – w młodości uznany włoski tenor operowy. Niestety, wskutek postępującej afazji utracił głos. Obecnie komunikuje się z pomocą syntezatora DEC Talk DTC, o czym śpiewa w melodramacie pt. „My voice and me”

Obszar badań eksperymentu

- Czym jest i jak się tworzy szczerość w głosie?
- Czy proces ten wygląda inaczej, gdy głos jest bezcielesny?
- Czy można wypełnić brak szczerości muzyką?

Nagranie utworu:

https://soundcloud.com/gravityisahat/by-hand-bounce-voice-favoured

Przebieg eksperymentu

Oliver Curry to fikcyjna postać, jego frazy wokalne zostały wygenerowane tylko i wyłącznie komputerowo. Partytura orkiestry została zagrana do podkładu solisty. Brak kontaktu fizycznego z wokalistą stanowił trudność dla muzyków.

Muzyka stanowiła ramę składniową dla utworu. Imitowała znaki interpunkcyjne oraz inne paralingwistyczne efekty.

Wyrazy dźwiękonaśladowcze (np. kasłanie, śmiech) miały za zadanie dodać dramaturgii, wzbudzić empatię i emocje w słuchaczach.

Test

1. ETI Eloquance, 1998
2. Polytehnic University of Madrid, 2010
3. autentyczny włoski
4. Acapela, 2013
5. Tokyo Institute of Technology Kobayashi Laboratory, 2004

Test

6. autentyczny japoński
7. MIT (DecTalk), 1989
8. IBM Watson Research Center, 2004
9. HAL „2001: a space odyssey”, 1968
10. Voxygen, 2012

Podsumowanie

- Dla percepcji istotne jest nie tylko to, co słyszymy (gesty, mimika, widza o rozmówcy)
- Dla autentyczności mowy ważna jest modulacja głosu, interpunkcja, akcent
- Człowiek podświadomie przypisuje intencje i emocje rozmówcy
- Kobiecy głos jest postrzegany jako bardziej szczery od męskiego
- Im dłuższy czas wypowiedzi tym bardziej wrażliwi jesteśmy na sztuczność mowy

Bibliografia

- R. Tadeusiewicz *„Sygnał mowy”*, 1988
- J. W. Mullennixa, St. E. Sterna, St. J. Wilsonc
- C. Dysonb „Social perception of male and female computer synthesized speech”
- P. Baker, Ch. Newell, G. Newell *„Can a computer-generated voice be sincere? A case study combining music and synthetic speech”*, 2013
- http://emosamples.syntheticspeech.de