

# Windows Search

apl@amu.edu.pl

## Streszczenie

Badano okna ważące ze względu na przydatność rezultatów analiz spektralnych dla pomiaru częstotliwości podstawowej  $F0$  i dla klasyfikacji dźwięków mowy. Stosowano dwa kryteria: wartość cepstrum mocy w punkcie  $T0$  (okres  $F0$ ) oraz wariancję składowych spektrum chwilowego jako miarę dynamiki spektrum. Dynamikę obliczano w odniesieniu do spektrum po ślepym rozplocie i liftracji. Badaniom poddano okna typu: Hamminga, Hanna, Keiser-Bessela, Blackmana-Harrisa, Blackmana-Nutalla, flat-top, Gaussa – 2 typy i okno prostokątne. Badania wykonano dla okien o szerokościach 64, 128, 256, 512 i 1024 punktów dla sygnału mowy próbkowanego z częstotliwością 16 kHz. Na materiał eksperymentalny składały się realizacje 6 samogłosek polskich. Pobrano w tym celu po 20 realizacji każdej samogłoski z wypowiedzi każdego osobnika (45) bazy danych Corpora. Przeanalizowano  $6 \times 20 \times 45 = 5400$  realizacji samogłosek wypowiedzianych głosami męskimi (29), żeńskimi (10) i dziecięcymi (6). Z każdej realizacji do analiz wybierano najlepszy odcinek. Kryterium tego wyboru była wartość cepstrum mocy w punkcie  $T0$  określonym z użyciem okna Hamminga.

Okazało się, że najlepsze są okna Gaussa z optymalnymi odchyleniami standardowymi. Drugim z kolei jest okno typu flat-top ze względu na dynamikę i okno Hanna ze względu na pomiar  $F0$ . Najgorsze jest okno prostokątne. Najlepszym kompromisem jest okno Blackmana-Harrisa. Wnioski wyciągnięto na podstawie testów post-hoc wykonanych z użyciem ANOVA powtarzanych pomiarów.

## Summary

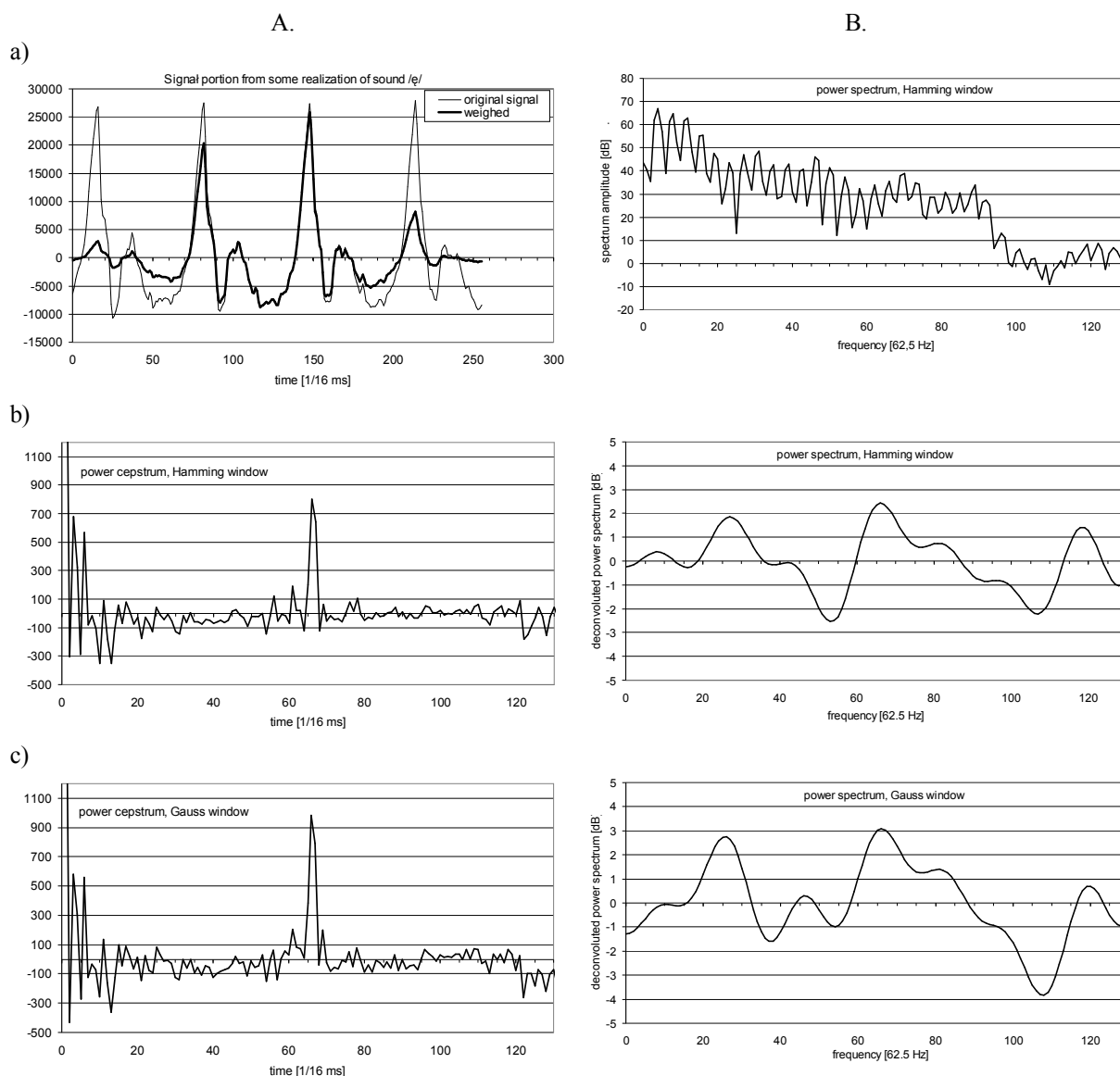
Signal weighting windows were examined for usefulness of results of spectral analysis for fundamental frequency ( $F0$ ) measurements and for speech sounds recognition. There were applied two criterion: the power cepstrum value at the  $T0$  point (the  $F0$  period) and the variance of of an instantaneous spectrum as a dynamics measure of the spectrum. The dynamics was computed with reference to the spectrum after blind deconvolution and liftration. There were examined nine windows types, i.e.: Hamming, Hann, Keiser-Bessel, Blackman-Harris, Blackman-Nuttall, flat-top, Gauss – 2 types and rectangular. We examined windows of five widths each: 64, 128, 256, 512 and 1024 points for a speech signal sampled at 16 kHz. The experimental material consisted of realizations of 6 polish vowels. There were taken for this purpose 20 realizations of each vowel from every speaker (45) from Corpora data basis. Thus there were analysed as much as  $6 \times 20 \times 45 = 5400$  vowels realizations spoken by male (29), female (10) and children (6) voices. There was chosen the best part from each vowel realization for the analysis. As a criterion of these choices was considered value of the power cepstrum at the  $T0$  point found using the Hamming window.

It turned out, that the best were Gauss windows with optimized standard deviation. The next, worse, were flat-top windows for the dynamics emphasize and Hann windows for the  $T0$  measurements. The worst turned out to be rectangular windows. The best compromise – a window suitable for both tasks were Blackmans windows. These conclusions were drawn on the basis of the ANOVA of repeated measures tests.

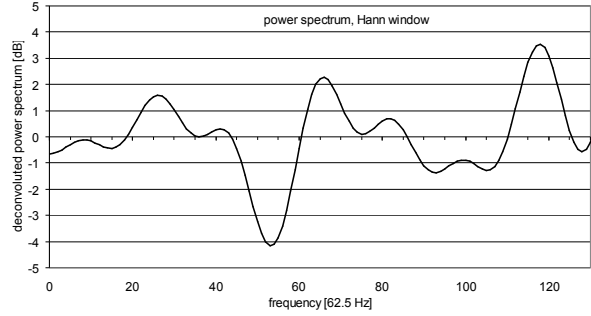
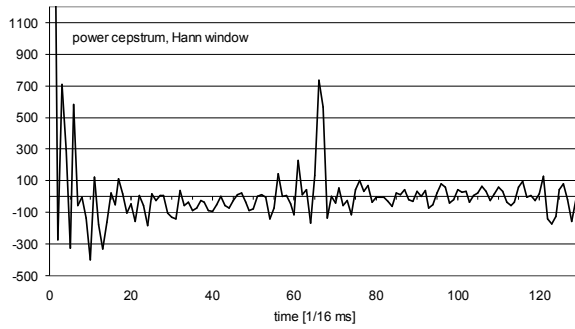
## Windows Search

**Geneza.** Na podstawie obserwacji wizualnych stwierdzono silną zależność postaci cepstrów mocy od wyboru okna ważącego. Zauważono, że zależność ta w szczególności dotyczy wartości cepstrum w punkcie  $T0$ , tj. w punkcie odpowiadającym okresowi podstawowemu. Spostrzeżenia te, zilustrowane rysunkami 1 – 3 w odniesieniu do badań pewnej jednej realizacji głoski /e/, stały się inspiracją do przebadania szerszego materiału eksperymentalnego. Szczególnej uwagi wymagało okno Gaussa, które ma jeden parametr swobodny.

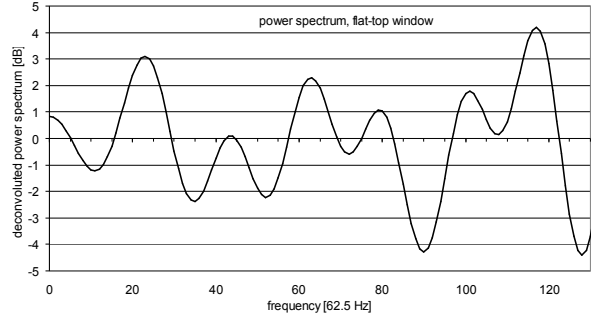
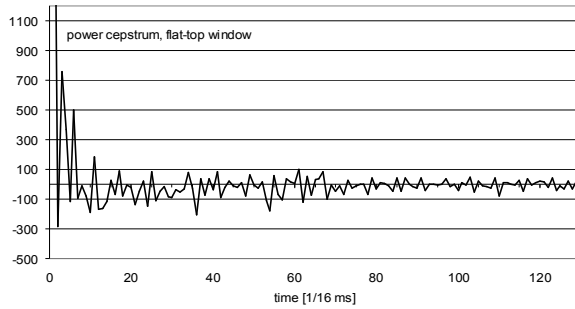
Mając na uwadze potrzeby związane z rozpoznawaniem (klasyfikacją) dźwięków obserwacje wartości cepstrum w punkcie  $T0$  połączono z obserwacjami dynamiki spektrum po ślepym rozplocie. Przyjęte kryteria badań uściślamy w następnym punkcie artykułu.



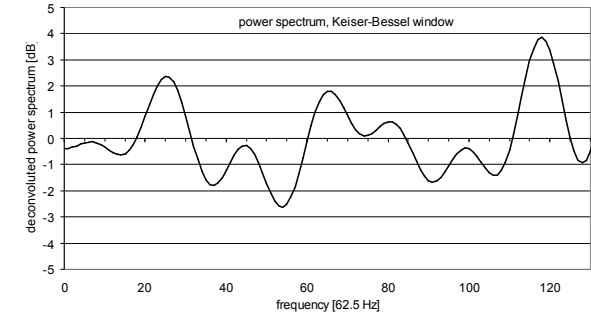
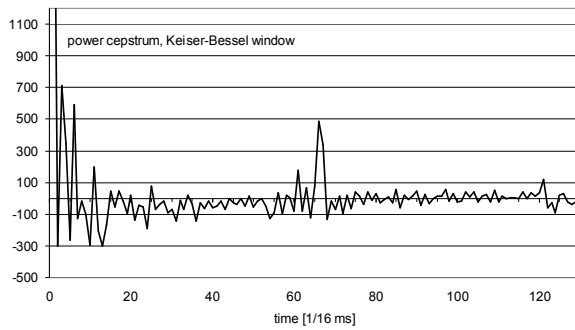
d)



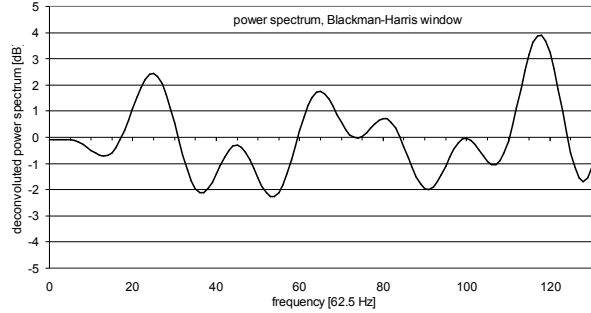
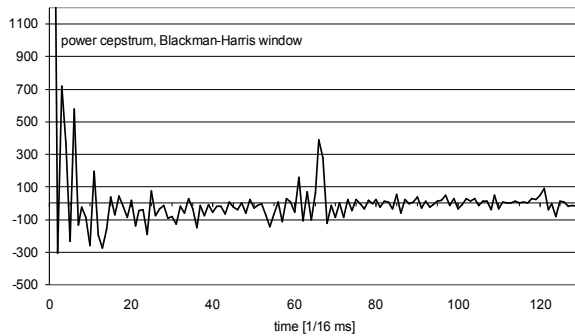
e)



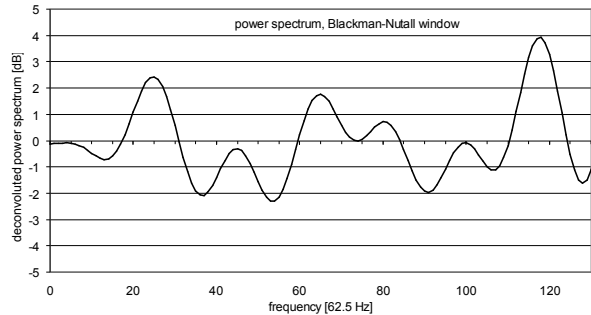
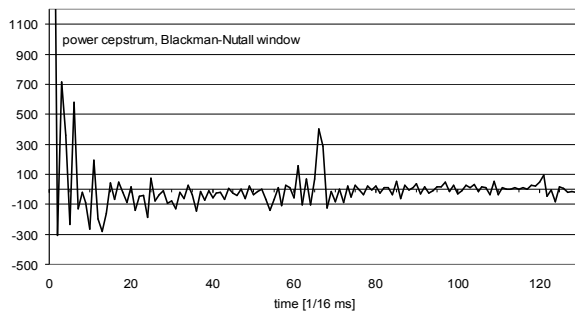
f)



g)



h)



i)

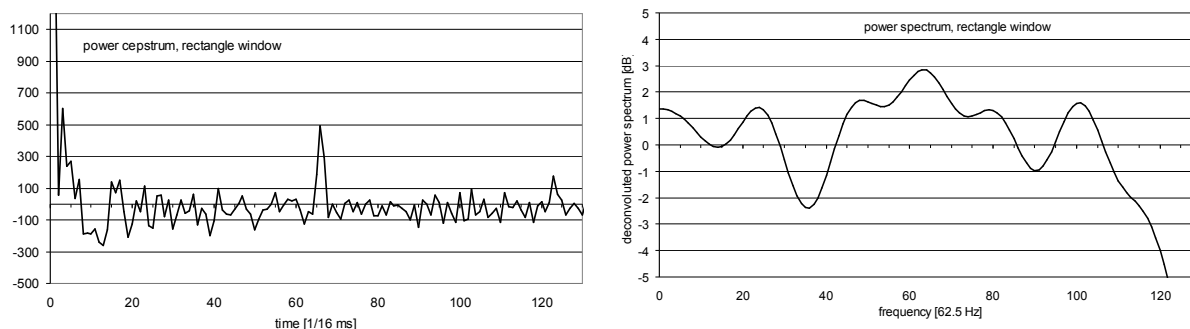


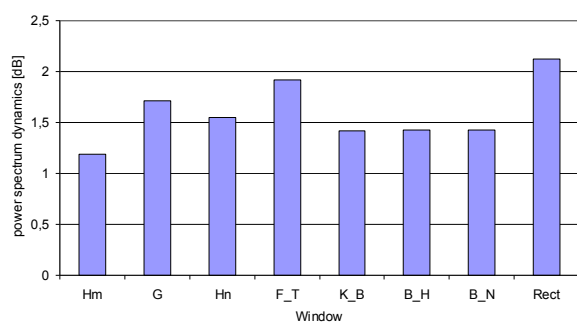
Fig. 1. Power cepstra (A.b-i) and spectra after blind deconvolution (B.b-i) of some realization of the sound /ę/ after applying different weighting windows; a – original signal and weighted with the Hamming window and its power spectrum after applying the Hamming window; windows width 256 pts, cepstrum(0) for deconvoluted spectrum was set to 0. Cepstrum drawings are clipped, so the cepstrum(0) and cepstrum(1) are not seen  
E:\apl\Spectrum\Windows\_Search\windows\_effects.xlsb

Table 1. Cepstrum at  $T0$  (=66 pts/16 pts/ms=4,125 ms) for different windows

E:\apl\Spectrum\Windows\_Search\windows\_effects.xlsb

window	cepstrum	dynamics
Hm	799,246	1,188
G	973,459	1,71
Hn	735,072	1,553
F_T	39,3	1,919
K_B	485,991	1,419
B_H	390,076	1,43
B_N	401,267	1,43
Rect	494,607	2,121

a)



b)

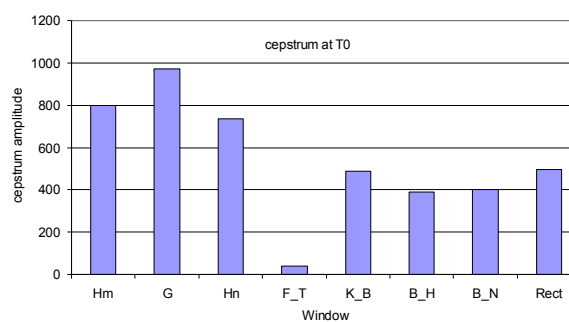


Fig. 2. Dynamika spektrum po ślepym rozplacie (a) i cepstra w punkcie  $T0$  (b) dla różnych okien ważących (szerokość okien – 256 punktów). E:\apl\Spectrum\Windows\_Search\windows\_effects.xlsb

Dynamics of the spectrum after blind deconvolution (a) and cepstra at the  $T0$  point (b) for different weighting windows (windows width – 256 points)

Hm=Hamming, G=Gauss, Hn=Hann, F\_T=flat-top, K\_B=Keiser-Bessel, B\_H=Blackman-Harris, B\_N=Blackman-Nutall, Rect=rectangle,  $T0$ =66 pts/16 pts/ms=4,125 ms

Table 2. Cepstrum at  $T0$  as function of the parameter  $\sigma$  of the Gauss's weighting window

E:\apl\Spectrum\Windows\_Search\windows\_effects.xlsb

$200\sigma$	$\sigma$	cepstrum at $T0$
10	0,050	0,000
20	0,100	28,000
25	0,125	10,000
30	0,150	30,000
35	0,175	100,000
40	0,200	160,000
50	0,250	300,000
60	0,300	410,000
70	0,350	550,000
80	0,400	690,000
85	0,425	760,000
90	0,450	842,000
95	0,475	931,000
97	0,485	968,880
100	0,500	982,530
102	0,510	973,450
105	0,525	967,800
110	0,550	963,700
115	0,575	975,390
120	0,600	976,690
125	0,625	978,550
130	0,650	985,200
135	0,675	995,700
137	0,685	1006,741
140	0,700	997,050
145	0,725	986,500
150	0,750	970,700
170	0,850	850,000
200	1,000	760,000
220	1,100	720,000
250	1,250	650,000
300	1,500	570,000
500	2,500	564,000
1500	7,500	558,000

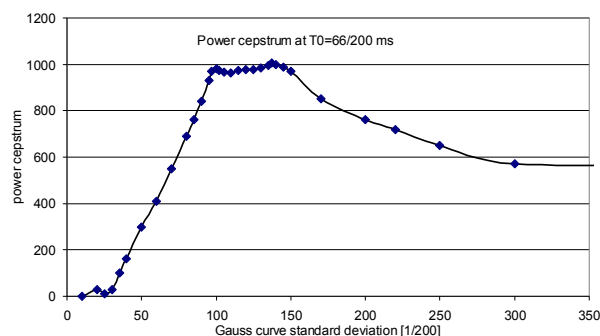


Fig. 3. Cepstrum values at  $T0$  for Gauss's windows for different parameters  $\sigma$ ; sound / $\epsilon$ /, windows width 256 pts  
E:\apl\Spectrum\Windows\_Search\windows\_effects.xlsb

**Badania – kryteria.** W analizie spektralnej stosuje się okna wążące aby pomniejszyć skutki nieciągłości na granicach wycinków sygnału. W literaturze można znaleźć informacje o wła-

snościach różnych okien ważących. Głównym kryterium ich oceny jest tłumienie wstęg bocznych spektrum okna. Spektrum dobrego okna ważącego powinno charakteryzować się wąskim pasmem w otoczeniu maksimum i możliwie najmniejszymi maksimami bocznymi. Najlepiej spełniają to kryterium okna typu "flat-top", gorzej okna Blackmana i okna Hamminga. Jest też kilka propozycji pośrednich [Randal].

W rozpoznawaniu dźwięków mowy, do wygładzania cepstralnego obwiedni spektrum oraz do pomiaru częstotliwości podstawowej pobudzenia (częstotliwości tonu krtaniowego –  $F0$ ) stosowane są analizy cepstralne. W tej metodzie pomiarowi podlega okres tonu podstawowego –  $T0$ . Pomiar  $T0$  będzie tym pewniejszy, im cepstrum w tym punkcie będzie mieć większą wartość. Niewątpliwie parametry okien ważących pozostają w związku ze specyfiką sygnału mowy. Specyfika ta wynika nie tylko z częstotliwościowego zakresu składowych spektralnych lecz także z szybkości artykulacji i z zakresu częstotliwości tonu krtaniowego – parametrów uwarunkowanych zdolnością percepcji i wymogiem pewności przekazywania informacji oraz parametrami toru głosowego człowieka, czyli z anatomii. W przeciętnych wypowiedziach artykułowanych jest nie więcej niż 15 głosek/s, a zakres częstotliwości tonu krtaniowego zawiera się w granicach od 70 do 800 Hz. Z tych właściwości sygnału mowy wynika, że próg liftracji nie powinien być wyższy niż 1,25 ms, a okres uśredniania w procedurach ślepego rozplotu nie powinien być krótszy niż  $1/15$  s ( $\approx 67$  ms). Widać, że ocieramy się tu o granice zakresów częstotliwości percepcji tonów prostych ( $>16$  Hz) i zakresów spektrum odpowiedzi impulsowej toru głosowego.

Opisane wyżej pobieżne obserwacje wskazują na kompletną nieprzydatność okien typu flat-top do pomiaru częstotliwości podstawowej. Cepstra wytworzone z użyciem tego okna przyjmowały w punkcie odpowiadającym położeniu tonu podstawowego małe, niekiedy nawet ujemne wartości (czyli wskazywały wręcz na brak tej składowej). W technologii sygnału mowy najważniejsze jednak jest rozpoznawanie, czyli klasyfikowanie dźwięków. Uznano więc, że należałoby też zbadać zależność tej zdolności od kształtu okna ważącego. Dźwięki mowy rozpoznaje się na podstawie kilku pierwszych składowych cepstrum (1. do 12.). Ze względu na łatwą interpretację fizyczną (maksima odpowiadają położeniu rezonansów przewodu głosowego) łatwiej jest oceniać cepstralnie wygładzone spektrum mocy. Liczy się tu dynamika zmienności – im dynamika będzie większa, tym lepiej będą zachowane informacje o różnicach pomiędzy dźwiękami. Jako kryterium użyteczności dla rozpoznawania (klasyfikacji) dźwięków przyjęliśmy więc wariancję wygładzonego spektrum mocy. Aby pomniejszyć zmienność pochodzącą z różnic międzyosobniczych i z wpływu akustyki pomieszczeń, poddaliśmy spektrum mocy procedurze ślepego rozplotu<sup>1</sup>.

Przydatność okien w zastosowaniu do analizy sygnału mowy będziemy więc oceniać na podstawie **dwóch kryteriów**: na podstawie wartości cepstrum w punkcie  $T0$  oraz na podstawie wielkości wariancji spektrum mocy poddanego ślepego rozplotaniu i wygładzeniu cepstralnemu. Aby zachować wymiar w dB posługujemy się pierwiastkiem z wariancji, czyli odchyleniem standardowym składowych spektralnych. Badaniu podlegało 8 rodzajów okien ważących: Hamminga, Hann, Gaussa, Keiser-Bessela, Blackman-Harrisa, Blackman-Nuttalla, flat-top i okno prostokątne. Wszystkie okna ważące były normalizowane poprzez dzielenie wyznaczanych przez nie wag przez wartość maksymalną, aby wartości wag zawierały się w przedziale  $(0,1)$ . Definicje matematyczne badanych okien podajemy w suplemencie 4.

**Okno Gaussa**, które ma jeden swobodny parametr – odchylenie standardowe – było badane dla kilkuset różnych jego wartości (200) celem znalezienia wartości optymalnej ze względu

<sup>1</sup> Ślepe rozplotanie polegało na odejmowaniu średniej bieżącej po czasie od aktualnego chwilowego spektrum mocy. Operacja ślepego rozplotu była wykonywana w dziedzinie cepstrum, które i tak musiało być obliczone z uwagi na stosowane wygładzenie cepstralne. W tym wypadku uśrednianie dziedzinie cepstrum obejmuje tylko składowe z pasma liftrowanego (u nas 14). W oparciu o pobieżną inspekcję wizualną skutków długości okresu uśredniania przyjęliśmy, jako okres uśredniania, czas 205 ms.

na wymienione wyżej kryteria. Jak napisano wyżej, z obserwacji efektów zastosowania różnych okien wynikało, że postać cepstrum silnie zależy od postaci okna ważącego. W szczególności dotyczy to obszaru determinowanego przez ton krtaniowy i mierzonej w efekcie częstotliwości podstawowej.

Zaobserwowaliśmy, na podstawie wizualnej inspekcji kilkunastu realizacji różnych samogłosek, że największe wartości cepstrum w punkcie  $T_0$  pojawiają się przy zastosowaniu okna Hamminga lub okna Gaussa dla  $\sigma \approx 0,5$ . Okno Hamminga jest najczęściej stosowane w analizach sygnału mowy. Detekcja  $T_0$  na podstawie cepstrum nie zawsze jest możliwa, np. przy użyciu okna typu flat-top. Z tego względu postanowiliśmy, obok metody wykrywania maksimum w każdym przypadku, posłużyć się pomiarami wartości cepstrum w punkcie  $T_0$  przy założeniu, że znamy prawdziwą wartość  $T_0$ . Zakładaliśmy, że częstotliwość podstawowa wyznaczana za pomocą okna Hamminga jest najbliższa owej "prawdziwej". Z tego względu, że zastosowanie go daje wysokie wartości cepstrum w punkcie  $T_0$ , postanowiliśmy użyć tego właśnie okna do określania  $T_0$  w każdej analizowanej realizacji dźwięku i odczytywać w tym punkcie wartości cepstrum dla wszystkich pozostałych badanych okien. Równolegle jednak odczytywano też wartości cepstrum w punkcie  $T_0$  znalezionym przy zastosowaniu bieżącego okna.

W zakres analiz spektralnych wchodziły obliczenia logarytmicznego spektrum mocy, cepstrum mocy, ślepego rozplotu w dziedzinie cepstrum i lifrowanego spektrum mocy (próg odcięcia – od 15. punktu) obliczanego na podstawie cepstrum po ślepym rozplotu (por. Fig. 1.B).

## Badania – wyniki

Celem badań było określenie przydatności różnych okien dla analizy sygnału mowy ze względu na wymienione kryteria. Do badań pobrano z wypowiedzi każdej osoby z bazy danych *Corpora* po 20 realizacji każdej z 6 samogłosek języka polskiego w mowie potocznej. Baza danych zawierała próbki wypowiedzi 45 osób różnej płci i wieku. Łącznie więc przebadano  $6 \times 20 \times 45 = 5400$  przykładów realizacji samogłosek..

Table 3. Optymalne odchylenia standardowe okien Gaussa w zależności od szerokości okna i zastosowań; zob. Supplement 1, Fig 1.1 oraz Fig. 1.2

Optimal standard deviation of the Gauss curve ( $\sigma$ ) for different width of the window and for different tasks, see Supplement 1, Fig 1.1 and Fig. 1.2

E:\apl\Spectrum\Windows\_Search\Windows\_Search\_64\_128\_256\_512\_1024\joined\_64\_128\_256\_512\_1024.xlsb

windows width	64	128	256	512	1024
recognition purpose	0,190	0,185	0,135	0,08	0,05
F0 measures purpose	0,45	0,465	0,505	0,405	0,27

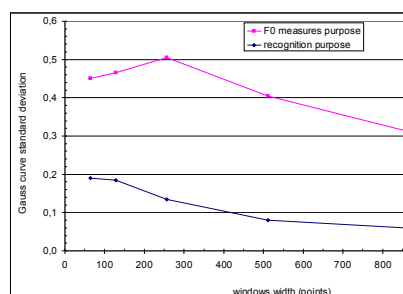


Fig. 4. Parameter "standard deviation for optimal Gauss Windows  
Parametr "odchylenie standardowe" optymalnych okien Gaussa

E:\apl\Spectrum\Windows\_Search\Windows\_Search\_64\_128\_256\_512\_1024\joined\_64\_128\_256\_512\_1024.xlsb

Badano, jak już wspomniano, użyteczność dziewięciu okien ważących, okna: Hamminga, Gaussa – dwa rodzaje, flat-top, Keisera-Bessela, Blackmana-Harrisa, Blackmana-Nutalla i prostokątnego. Zasadnicze wyniki badań prezentujemy poniżej w zestawieniach tabelarycznych (tabele 3 – 5) i za pomocą odpowiednich wykresów (Fig. 4. – 6.). W opisie wyników



podajemy nazwy zmiennych reprezentujących badane cechy zdarzeń w programie komputerowym i w raportach generowanych przez ten program, co ułatwi Czytelnikowi ich czytanie. Jakościowe podsumowanie wyników badań przedstawiono w tabeli 6.

Table 4. Dynamika spektrum po ślepym rozplocie<sup>2</sup>. Wartości średnie i odchylenia standardowe dynamiki  
Dynamics of spectrum after blind deconvolution. Mean values and standard deviations of dynamics

E:\apl\Spectrum\Windows\_Search\Windows\_Search\_64\_128\_256\_512\_1024\joined\_64\_128\_256\_512\_1024.xlsb

		Windows name								
Windows width	What	Ham-ming	Gauss-StDev	Gauss T0	Hann	Flat-top	Keiser-Bessel	Blackman-Harris	Blackman-Nutal	Recta-nle
64	mean value	<b>7,69</b>	<b>7,88</b>	<b>7,49</b>	<b>8,15</b>	<b>7,91</b>	<b>8,25</b>	<b>8,20</b>	<b>8,21</b>	<b>5,33</b>
	st. dev.	2,66	2,91	2,65	2,74	2,89	2,79	2,81	2,81	6,57
128	mean value	<b>7,21</b>	<b>8,32</b>	<b>7,03</b>	<b>7,72</b>	<b>8,23</b>	<b>7,93</b>	<b>8,03</b>	<b>8,02</b>	<b>5,06</b>
	st. dev.	2,62	2,79	2,62	2,65	2,76	2,67	2,69	2,69	2,11
256	mean value	<b>7,09</b>	<b>8,17</b>	<b>6,76</b>	<b>7,40</b>	<b>7,86</b>	<b>7,54</b>	<b>7,60</b>	<b>7,60</b>	<b>5,41</b>
	st. dev.	2,64	2,77	2,62	2,70	2,75	2,74	2,76	2,76	2,68
512	mean value	<b>6,50</b>	<b>8,08</b>	<b>6,67</b>	<b>6,79</b>	<b>7,56</b>	<b>7,02</b>	<b>7,16</b>	<b>7,14</b>	<b>5,33</b>
	st. dev.	2,43	2,80	2,48	2,49	2,74	2,57	2,63	2,62	2,27
1024	mean value	<b>6,33</b>	<b>8,01</b>	<b>6,63</b>	<b>6,50</b>	<b>6,89</b>	<b>6,55</b>	<b>6,60</b>	<b>6,59</b>	<b>5,33</b>
	st. dev.	2,36	2,80	2,38	2,37	2,49	2,35	2,37	2,36	2,32

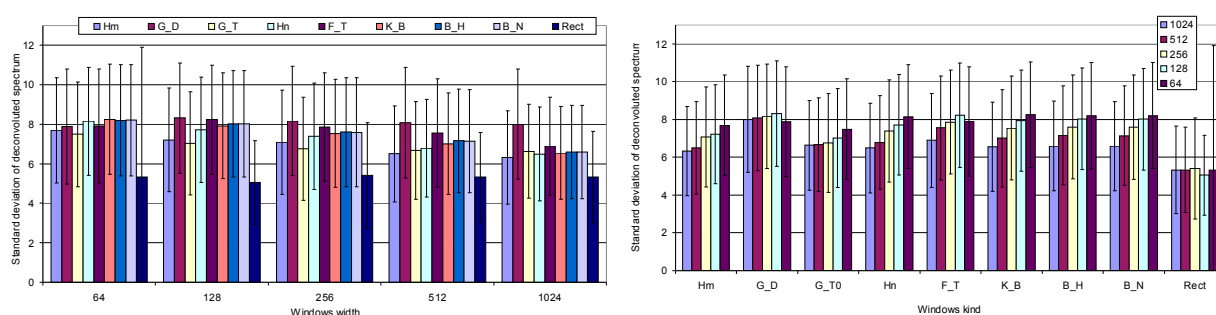


Fig. 5. Dynamika spektrów w zależności od rodzaju okna ważącego. Słupki błędów wyrażają wartość odchylenia standardowego dynamiki; G\_D – okno Gaussa optymalizowane ze względu na dynamikę spektrum po ślepym rozplocie, G\_T – okno Gaussa optymalizowane ze względu na pomiar T0

Blind deconvoluted spectrum dynamics as a dependence of weighting Windows kind. Errors marks are for standard deviations; G\_D – Gauss window optimized for spectrum dynamics, G\_T – Gauss window optimized for T0 measurements

E:\apl\Spectrum\Windows\_Search\Windows\_Search\_64\_128\_256\_512\_1024\joined\_64\_128\_256\_512\_1024.xlsb

Table 5.A. Wartości średnie i odchylenia standardowe cepstrum w punkcie T0<sup>3</sup> znalezionym za pomocą okna Hamminga

Mean values and standard deviations of cepstrum at T0 point found using Hamming windows

E:\apl\Spectrum\Windows\_Search\Windows\_Search\_64\_128\_256\_512\_1024\joined\_64\_128\_256\_512\_1024.xlsb

		Windows name								
Windows width	What	Ham-ming	Gauss-StDev	Gauss T0	Hann	Flat-top	Keiser-Bessel	Blackman-Harris	Blackman-Nutall	Rectangle
64	mean	115	2	108	55	-8	34	22	24	3
	st. dev.	23	22	28	48	25	41	37	38	84
128	mean	213	42	208	164	28	133	111	114	63
	st. dev.	55	74	65	88	84	89	93	92	100
256	mean	635	53	637	582	53	425	339	349	298
	st. dev.	345	128	334	347	188	284	248	251	212
512	mean	1745	66	1740	1756	784	1624	1480	1497	653
	st. dev.	686	174	716	732	840	824	844	842	342
1024	mean	2784	253	3269	2942	3027	3198	3262	3258	1165
	st. dev.	1148	497	1428	1245	1599	1343	1398	1392	681

<sup>2</sup> zmienna *vPSstDevMaxGaussValue*

<sup>3</sup> zmienna *cpstr\_T0fixedAmplitude*



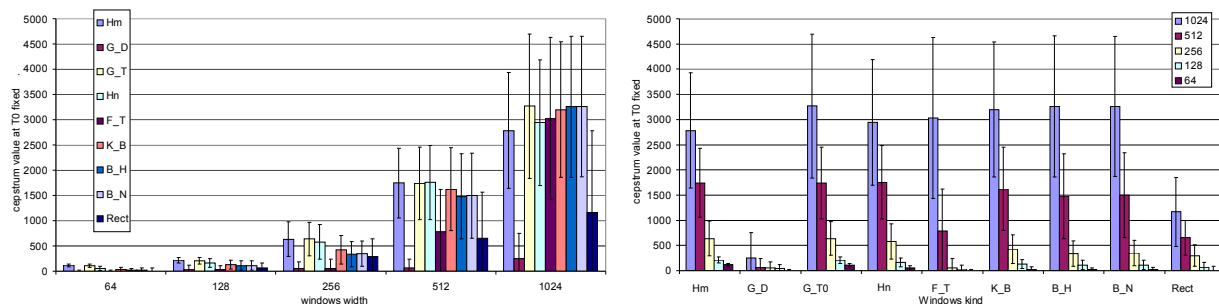


Fig. 6.A. Wartości cepstrum w punkcie  $T_0$  znalezionym za pomocą okna Hamminga w zależności od rodzaju okna ważącego. Słupki błędów wyrażają wartość odchylenia standardowego  
Cepstrum values at  $T_0$  point found with the Hamming window as a dependence of weighting windows kind.  
Error bars show standard deviations values

E:\apl\Spectrum\Windows\_Search\Windows\_Search\_64\_128\_256\_512\_1024\joined\_64\_128\_256\_512\_1024.xlsb

Duże wariancje (uwidocznione na wykresach jako odchylenia standardowe) sugerują nieistotność różnic związaną z rodzajem okna ważącego. Tak jednak nie musi być, ponieważ wartości cepstrów w funkcji rodzaju okna ważącego mogą być silnie dodatnio skorelowane. Wtedy nie zdarzy się, że cepstra pewnego podzbioru dźwięków będą rosnać przy zastosowaniu pewnego okna, a cepstra innego podzbioru spadać przy tym, a zmienność manifestująca się wariancją uwidoczną na wykresach będzie wynikać ze zróżnicowania cech wartości badanych realizacji dźwięków, a nie ze zróżnicowania okien. Wyniki obliczeń zawarte w tabelach suplementu 3. ujawniają istotne skorelowanie wartości cech określanych z użyciem różnych okien, a testy ANOVA powtarzanych pomiarów (*repeated measures*) wykazują w niemal wszystkich przypadkach bardzo istotne różnice.

Tabela 5.B. Wartości średnie i odchylenia standardowe cepstrum w punkcie  $T_0$ <sup>4</sup>

Mean values and standard deviations of cepstrum at  $T_0$  point found using current windows

E:\apl\Spectrum\Windows\_Search\Windows\_Search\_64\_128\_256\_512\_1024\joined\_64\_128\_256\_512\_1024.xlsb

Windows name										
Win-dows width	What	Ham-ming	Gauss StDev	Gauss T0	Hann	Flat_Top	Keiser-Bessel	Black man-Harris	Black man-Nutal	Rect angle
64	mean	115	33	110	80	41	63	55	56	41
	st. dev.	23,5	19,1	27,2	31,8	19,8	27,6	25,3	25,8	81,6
128	mean	213	125	209	186	127	168	161	162	118
	st. dev.	54,8	52,2	64,4	66,0	51,4	65,6	65,1	65,4	80,4
256	mean	635	270	639	591	273	449	387	393	334
	st. dev.	345	116	332	337	118	263	213	218	185
512	mean	1745	487	1740	1759	930	1630	1491	1507	692
	st. dev.	686	215	715	727	732	817	832	830	312
1024	mean	2784	1007	3311	2946	3143	3215	3294	3287	1277
	st. dev.	1148	432	1394	1240	1508	1328	1371	1366	628

**Correlation and ANOVA analysis results.** Celem weryfikacji istotności różnic ocen badanych cech przy użyciu różnych okien przeprowadzono analizy korelacji tych ocen oraz analizy wariancji powtarzanych pomiarów, gdzie zmienną był rodzaj okna ważącego<sup>5</sup>. Celem oceny istotności wpływu wyboru okna na pomiar  $T_0$  poddano analizie statystycznej dwa parametry

<sup>4</sup> zmienna  $cpstr\_T0currentAmplitude$

<sup>5</sup> W tych analizach dźwięki, czyli elementy populacji, były widziane w jako niezależne indywidua ("osobniki"), zaś aplikowanie poszczególnych okien ważących jako "działanie czasowe".

try: wartości  $T0$  pomierzone z użyciem badanych okien i wyniki pomiarów cepstrum dla każdego z tych okien w punkcie  $T0$  oraz w punkcie  $T0$  odczytanym dla okna Hamminga. Analizom statystycznym poddano też wyniki obliczeń dynamiki spektrum po ślepym rozplocie. Pełne wyniki obliczeń współczynnika korelacji liniowej i testów post-hoc przytaczamy w dodatku (suplement 3). Częściowe wyniki badania korelacji (średnie) zamieszczamy w tabelach i rysunkach w suplemencie 2.

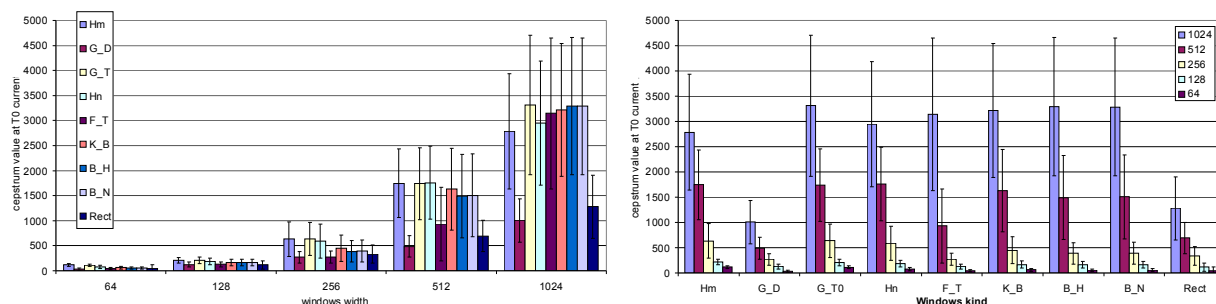


Fig. 6. B. Wartości cepstrum w punkcie  $T0$  w zależności od rodzaju okna ważącego. Słupki błędów wyrażają wartość odchylenia standardowego

Cepstrum values at the  $T0$  point as a dependence of weighting windows kind. Error bars show standard deviation  
E:\apl\Spectrum\Windows\_Search\Windows\_Search\_64\_128\_256\_512\_1024\joined\_64\_128\_256\_512\_1024.xlsb

**Statystyki dla okresu podstawowego<sup>10</sup>.** W tabeli 6. pokazano uszeregowanie rodzajów okien wg testów post-hoc – od najgorszych do najlepszych (wg rosnących wartości wskazanej cechy). ANOVA pomiarów powtarzanych wykazuje istotne statystycznie zróżnicowanie odczytów  $T0$  w zależności od rodzaju okna ważącego przy wszystkich badanych szerokościach. Wszystkie okna **256** punktowe różnią się pod tym względem istotnie. Hipoteza zerowa o równości średnich jest w większości przypadków odrzucana przy poziomie istotności  $p \approx 0$  (program Statistica podaje wartość 0) za wyjątkiem 2 przypadków – okna pary (B\_H, B\_N) dają wyniki istotnie różne na poziomie istotności  $p=0,009$ , a pary (Hm, G\_T) na poziomie istotności  $p=0,001$ . Przy oknach o szerokościach **512** – pojawiają się 3 grupy jednorodne.

Table 6. Uszeregowanie okien i grupy jednorodne wg testów post-hoc, NIR

Windows ranking and homogenous groups according to LSD post-hoc tests

E:\apl\Spectrum\Windows\_Search\Windows\_Search\_64\_128\_256\_512\_1024\Windows\_Ranking.xlsb

windows width [points]																			
$T0^6$					cepstrum "fixe $T0^7$					cepstrum "current $T0^8$					dynamics <sup>9</sup>				
64	128	256	512	1024	64	128	256	512	1024	64	128	256	512	1024	64	128	256	512	1024
G_D	G_D	G_D	G_D	G_D	F_T	F_T	G_D	G_D	G_D	G_D	Rect	G_D	G_D	G_D	Rect	Rect	Rect	Rect	Rect
F_T	F_T	F_T	F_T	F_T	G_D	G_D	F_T	Rect	Rect	Rect	G_D	F_T	Rect	Rect	G_T	G_T	G_T	Hm	Hm
B_H	B_H	B_H	B_H	G_T	Rect	Rect	Rect	F_T	Hm	F_T	F_T	Rect	F_T	Hm	Hm	Hm	Hm	G_T	Hn
B_N	B_N	B_N	B_N	B_H	B_H	B_H	B_H	B_H	Hn	B_H	B_H	B_H	B_H	Hn	G_D	Hn	Hn	Hn	K_B
K_B	K_B	K_B	K_B	B_N	B_N	B_N	B_N	B_N	F_T	B_N	B_N	B_N	B_N	F_T	F_T	K_B	K_B	K_B	B_N
Hn	Hn	Hn	Hn	K_B	K_B	K_B	K_B	K_B	K_B	K_B	K_B	K_B	K_B	K_B	Hn	B_N	B_N	B_N	B_H
Hm	Hm	Hm	G_T	Hn	Hn	Hn	Hn	G_T	B_N	Hn	Hn	Hn	G_T	B_N	B_H	B_H	B_H	B_H	G_T
G_T	G_T	G_T	Hm	Hm	G_T	G_T	Hm	Hm	B_H	G_T	G_T	Hm	Hm	B_H	B_N	F_T	F_T	F_T	F_T
Rect	Rect	Rect	Rect	Rect	Hm	Hm	G_T	Hn	G_T	Hm	Hm	G_T	Hn	G_T	K_B	G_D	G_D	G_D	G_D

<sup>6</sup> zmienna  $T0_{currentIndex}$

<sup>7</sup> zmienna  $cpstr\_T0_{fixedAmplitude}$

<sup>8</sup> zmienna  $cpstr\_T0_{currentAmplitude}$

<sup>9</sup> zmienna  $vPSstDevValue$

<sup>10</sup> zmienna  $T0_{currentIndex}$

Wyniki uzyskane tu za pomocą okien  $B\_H$  i  $B\_N$  nie różnią się istotnie przy poziomie istotności 0,212, a okno  $G\_T$  wchodzi w skład 2 grup jednorodnych  $\{Hn, G\_T, Hm\}$  i  $\{G\_T, Hm, Rect\}$ . Okno  $G\_T$  nie daje istotnie różnych odczytów  $T0$  od odczytów okna  $Hn$  przy  $p=0,331$ , i od okna  $Hm$  przy  $p=0,666$ , a od okna  $Rect$  przy  $p=0,230$ .

Przy oknach o szerokości **1024** testy post-hoc ujawniają się 4 mocno pokrywające się grupy jednorodne skupiające po 4 okna:

$\{G\_T, B\_H, B\_N, K\_B\}$ ,  
 $\{B\_H, B\_N, K\_B, Hn\}$ ,  
 $\{B\_N, K\_B, Hn, Hm\}$  i  
 $\{K\_B, Hn, Hm, Rect\}$ .

Pokazują one w szczególności, że odczyty  $T0$  z użyciem okna  $G\_T$  nie różnią się istotnie od odczytów  $T0$  z użyciem okna  $B\_H$  ( $p=0,605$ ),  $B\_N$  ( $p=0,579$ ) i  $K\_B$  ( $p=0,170$ ).

Grupy jednorodne okien  $\{G\_T, Hm\}$  pojawiają się także przy szerokościach **64** ( $p=0,671$ ) i **128** ( $p=0,314$ ) punktów. Wyniki te podajemy pro forma, okna te są zbyt wąskie dla pomiarów  $T0$ .

**Wnioski.** Oczekiwano niezależności wyników pomiaru  $T0$  w zależności od użytego okna, co powinno przejawiać się brakiem podstaw do odrzucenia hipotezy o równości średnich w teście ANOVA i w konsekwencji totalną homogenicznością w testach post-hoc. Większe grupy homogeniczne pojawiają się jednak dopiero przy oknach 512 punktowych i szerszych (1024 punktowych). Wyniki te pokazują w szczególności, że okna węższe niż 256 punktów nie nadają się do pomiarów  $T0$  w sygnale mowy (przy częstotliwości próbkowania 16 kHz), gdyż średnie położenie  $T0$  osiąga 100. punkt<sup>11</sup>, a te okna oferują maksymalnie tylko 32 i 64 punkty odpowiednio.

Wartość  $T0$  nie stanowi kryterium do rankingu okien – liczy się homogeniczność grup. Niższe wartości średnie  $T0$  dla okien  $G\_D$  i  $F\_T$  wynikają z częstego wchodzenia w obszar odpowiedzi impulsowej podczas poszukiwania maksimum w dziedzinie cepstrum mocy. Uszeregowanie okien wg wartości  $T0$  świadczy raczej tylko o średniej różnicy pomiędzy wynikami pomiarów z użyciem kolejnych okien.

**Statystyki wartości cepstrum w punkcie  $T0$  odczytanym dla okna Hamminga<sup>12</sup>.** Testy post-hoc wykazują tu nieliczne grupy jednorodne okien – jedną lub dwie w zależności od okna ważącego.

W kategorii okien o szerokości **256** są dwie grupy jednorodne

$\{Hm, G\_T\}_{p=0,638}$  i  
 $\{G\_D, F\_T\}_{p=0,856}$ .

W kategorii okien o szerokości **512** ujawniają się dwie częściowo pokrywające się grupy jednorodne

$\{G\_T, Hm\}_{p=0,478}$  i  
 $\{Hm, Hn\}_{p=0,105}$ .

W kategorii okien o szerokości **1024** ujawnia się grupa

$\{B\_N, B\_H, GT\}$  przy  $p=0,660$  dla pary  $(B\_N, B\_H)$ ,  $p=0,517$  dla pary  $(B\_H, G\_T)$  i  $p=0,276$  dla pary  $(G\_T, B\_N)$ .

W grupie okien **64** punktowych ujawnia się tylko jedna grupa jednorodna okien  $\{G\_D, Rect\}$ , a w grupie okien **128** punktowych nie ma grup jednorodnych.

**Statystyki wartości cepstrum w punkcie  $T0$  odczytanym dla bieżącego okna<sup>13</sup>.** Do grup homogenicznych okien ze względu na pomiary wartości cepstrum w punkcie  $T0$  dochodzą tu

<sup>11</sup> i.e.  $100/rate=100/16000=6,25$  ms, tj. 160Hz

<sup>12</sup> zmienna *cpstrT0fixedAmplitude*

<sup>13</sup> zmienna *castr\_T0currentAmplitude*

zbiory  $\{B\_H, B\_N\}$  dla wszystkich szerokości za wyjątkiem 512, gdzie wyniki dla tej pary różnią się. Poziom istotności, to  $p=0,202, 0,353, 0,063, 0,03$  i  $0,556$  dla okien tej pary o szerokościach od 64 do 1024 punktów odpowiednio, podczas gdy przy pomiarach cespstrum w punkcie  $T0$  określonym dla okna Hamminga wyniki uzyskiwane przy użyciu tych okien były klasyfikowane jako istotnie różniące się ( $p \leq 0,021$ ). Przy szerokości 512 punktów różnią się też te cespstra dla okien pary  $\{H_m, H_n\}$  na poziomie istotności  $p=0,05$ .

**Wnioski.** Wobec istotnych różnic wartości  $T0$  w zależności od zastosowanego okna okazało się, że nie do utrzymania jest założenie, że  $T0$  wskazane za pomocą okna Hamminga wskazuje "prawdziwą" wartość tego parametru. Należało więc poszerzyć badania o analizy statystyczne cespstrum w punkcie  $T0$  odczytanym z użyciem badanego okna, a nie odczytywać zawsze cespstrum w punkcie  $T0$  odczytanym za pomocą okna Hamminga. Podejście to, jak się okazało, podwyższyło wartości cespstrów, nie zmieniając jednak zasadniczego uszeregowania użyteczności okien ze względu na pomiar  $T0$  – por. tabele wyników dla zmiennych *cpstr\_T0FixedAmplitude* i *cpstr\_T0currentAmplitude*, suplement 3, tabele 5.A., 5.B. i rys. 6.A. i 6.B.

**Statystyki dynamiki**<sup>14</sup>. Testy post-hoc nie ujawniają grup jednorodnych okien ze względu na tę cechę za wyjątkiem bardzo zbliżonych parametrami okien  $B\_H$  i  $B\_N$  we wszystkich badanych szerokościach ( $p=0,829, 0,549, 0,707, 0,292$  i  $0,633$  dla badanych szerokości odpowiednio) oraz zbiorami

$\{G\_D, F\_T\}$ ,  $\{H_n, B\_H, B\_N\}$  i  $\{B\_H, B\_N, K\_B\}$

ujawniającymi się przy zastosowaniu nie mających znaczenia dla praktyki okien 64 punktowych.

**Wnioski.** Zgodnie z oczekiwaniami oknami zapewniającymi najlepszą rozpoznawalność będą okna ważące typu  $G\_D$  i  $F\_T$  a także na odleglejszych pozycjach – okna typu  $B\_H$  i  $B\_N$ . Najgorzej pod tym względem przedstawiają się okna prostokątne.

## Parametry analiz

Badania wykonano przy następujących parametrach analiz

Table 7. Parameters of analysis

Windows widths	64, 128, 256, 512, 1024
Cepstral lifter cut-of threshold	15
Nbr of frames in averaging <sup>15</sup>	41
Nbr of sounds examples	20
Range of Gauss windows standard deviation searching	$\langle 1, 200 \rangle / 200$ z krokiem co 1
Best part searching range	-100, 100 ms

Jednym z istotnych parametrów jest zakres poszukiwań najlepszych wycinków sygnału do analiz. Przyjęto tu przedział przeszukiwania  $\pm 100$  ms. Wybór ten zweryfikowano badając histogramy odchyłek położenia najlepszych wycinków od centrów realizacji badanych realizacji dźwięków. Histogramy te pokazują (Fig. 7., tabela 8.), że zakres poszukiwań najlepszych wycinków był wystarczająco szeroki. Tylko w 3 przypadkach na 5400 przy analizach oknem 128 punktowym i w 1 przypadku przy analizach oknem 512 punktowym przeszukiwanie dałoby być może wybór jeszcze lepszego odcinka przy przesunięciach bardziej w lewo od centrum. Skrajne odległości były raczej wynikiem błędów pomiarów  $T0$ , jak to zdają się potwierdzać histogramy wartości  $T0$ .

<sup>14</sup> zmienna *vPSstDevValue*

<sup>15</sup> for the blind deconvolution process

Table 8. Liczebności odchylek położenia najlepszych wycinków realizacji dźwięków od środka realizacji  
Counts of deviations of the best portions values from a sound center  
E:\apl\Spectrum\Windows\_Search\Histograms of deviations.xlsb

deviation [5 ms]	windows width				
	64	128	256	512	1024
	count of deviations	count of deviations	count of deviations	count of deviations	count of deviations
-20	0	3	0	1	0
-19	0	0	0	0	0
-18	5	0	0	0	0
-17	8	2	2	0	0
-16	4	6	1	2	1
-15	9	8	0	2	0
-14	28	12	2	0	2
-13	36	20	5	2	2
-12	44	31	11	11	4
-11	80	68	17	4	7
-10	87	82	20	13	11
-9	125	111	43	19	17
-8	194	174	67	32	22
-7	236	235	96	60	36
-6	275	274	165	99	72
-5	294	278	206	152	103
-4	299	315	282	236	150
-3	296	314	332	357	275
-2	313	311	447	475	460
-1	308	310	504	570	672
0	307	264	505	720	1419
1	320	336	554	690	851
2	318	373	549	652	569
3	291	324	436	448	298
4	319	327	341	297	184
5	312	306	281	238	124
6	257	261	185	147	56
7	221	218	114	76	28
8	134	153	74	52	11
9	100	104	66	24	15
10	76	64	40	11	5
11	45	45	27	6	5
12	27	30	13	2	1
13	15	16	4	1	0
14	5	11	4	1	0
15	7	7	1	0	0
16	2	4	2	0	0
17	1	3	2	0	0
18	2	0	2	0	0
19	0	0	0	0	0
20	0	0	0	0	0

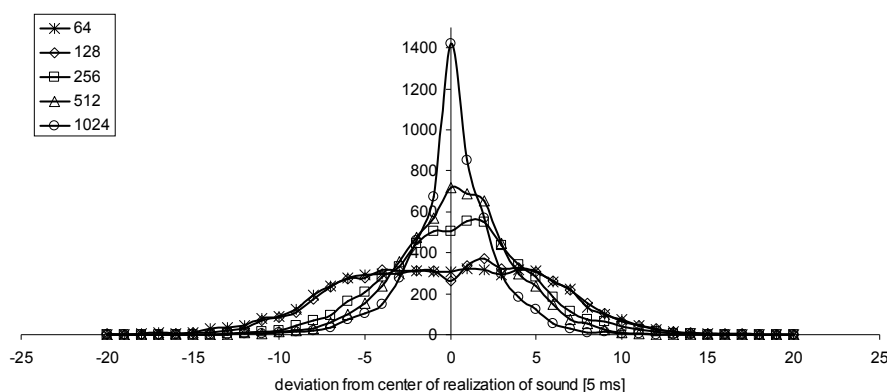


Fig. 7. Histogramy odchyłek położenia najlepszych wycinków realizacji dźwięków od środków realizacji (the Hamming window was applied)

Histogramm of values of the best portions deviations from a sound's centers

E:\apl\Spectrum\Windows\_Search\Histograms of deviations.xlsb

## Wyniki i dyskusja

Uiszeregowanie okien oraz ich grupy jednorodne ze względu na badane cechy zdarzeń w sygnale mowy zostało pokazane w tabeli 6. Poniżej przedstawiamy końcowe wnioski płynące z wykonanych badań.

**Klasyfikacja dźwięków mowy.** Zamieszczone wyniki pokazują, że rodzaj okna może mieć istotny wpływ na zdolności do klasyfikacji dźwięków mowy, dotyczy to w szczególności okna prostokątnego, dla którego dynamika spektrów ślepo rozplatanych jest wyraźnie niższa od pozostałych (rys. 5., tabela 4.). Najlepsze predyspozycje do tego celu ma okno Gaussa przy odpowiednio dobranym odchyleniu standardowym jako jego parametru. Kolejne miejsca zajmują okna typu flat-top, Blackmana-Harrisa i Blackmana\_Nutalla (tabela 6.). Najgorsze, to okna prostokątne, Gaussa optymalizowane dla pomiarów  $T0$  i Hamminga. ANOVA pokazuje istotne różnice, za wyjątkiem okien Blackmana, ale rozstrzygające byłoby badanie wyników analizy skupień – spodziewamy się tu wykryć większą homogeniczność właściwości okien z punktu widzenia zastosowań. Homogeniczność grup okien będzie bowiem zależeć od żądanej liczby skupień zdarzeń dla celów budowy książki kodów. W analizie skupień początkowo wszystkie zdarzenia będą zaliczane do jednej grupy niezależnie od rodzaju użytego okna. W miarę wzrostu liczby grup będą pojawiać się rozbieżności w klasyfikacji zdarzeń według parametrów obliczonych w efekcie stosowania różnych okien ważących i w konsekwencji będziemy mieć coraz mniejsze grupy homogeniczne okien. Wcześniejsze badania zależności promienia informacji od liczby skupień pokazały, że do rozróżnienia polskich fonemów powinno być wyróżnionych co najmniej 150 grup zdarzeń (zob.

<http://main3.amu.edu.pl/~apl/LVQ/Paper.zip>). Te badania zamierzamy wykonać.

**Pomiar  $T0$ .** Analizując średnie wnosimy, że najlepszymi właściwościami dla celów pomiaru  $T0$  charakteryzuje się optymalizowane pod tym kątem okno Gaussa, a zupełnie nie nadają się do tego celu ani okno Gaussa optymalizowane pod kątem klasyfikacji dźwięków, ani okno flat-top (Fig. 6., tabela 5.). Wyniki porównywalne do wyników dedykowanego okna Gaussa daje tu okno Hamminga. Odczyty cepstrum wskazują, że również okno Kaisera-Bessela i okno Blackmana-Harrisa powinny dać dobre wyniki w zastosowaniu do pomiaru  $T0$ .

**Analiza odchyłek pomiarów  $T0$ .** Ostatecznym kryterium poprawności wyboru okna ważącego ze względu na pomiar  $T0$  będzie rozbieżność wyników powstałych z zastosowania określonego okna ważącego. Określano rozbieżność wyników pomiarów  $T0$  dla par okien oraz tę rozbieżność w odniesieniu do okna Hamminga – to, aby ocenić wpływ błędów tu popełnianych na wybór najlepszych wycinków realizacji dźwięków. Rozbieżność mierzono pierwiast-

kiem kwadratowym ze średniej kwadratów odchyłek dla różnych okien względem każdego innego okna oraz, ze względu na wyróżnioną rolę tego okna, tylko względem okna Hamminga. Wyniki pokazują Fig. 8., 9., i tabela 9 oraz Fig. 10., 11., i tabela 10. W tabeli 9. zamieszczono średnie kwadratowe odchyłek odczytów  $T0$  i  $F0$  w odniesieniu do odczytów dokonanych za pomocą okna Hamminga, a w tabeli 10. i odpowiadających jej rysunkach odchyłki uśrednione po drugich oknach. Obliczenia wykonano posługując się wzorami

$$(1) \quad \overline{\Delta T_{i,j}} = \sqrt{\frac{1}{E} \sum_{e=1}^E (T_{i,e} - T_{j,e})^2} \quad \text{dla macierzy odchyłek } T0 \text{ oraz}$$

$$(2) \quad \overline{\overline{\Delta T_i}} = \frac{1}{W} \sum_{j=B\_H}^{Rect} \sqrt{\frac{1}{E} \sum_{e=1}^E (T_{i,e} - T_{j,e})^2} \quad \text{dla średnich różnic po oknach względem wybranego okna,}$$

gdzie  $e$  oznacza numer przykładu,  $E$  oznacza liczbę przykładów,  $j$  oznacza rodzaj okna,  $W$  oznacza liczbę badanych okien,  $T[s]$  oznacza okres tonu podstawowego.

Odchyłki częstotliwości podstawowej  $F0[Hz]$  obliczano analogicznie wg wzorów

$$(3) \quad \overline{\Delta F_{i,j}} = \sqrt{\frac{1}{E} \sum_{e=1}^E \left( \frac{T_{i,e} - T_{j,e}}{T_{i,e} T_{j,e}} \right)^2} \quad \text{dla macierzy odchyłek } F0 \text{ oraz}$$

$$(4) \quad \overline{\overline{\Delta F_i}} = \frac{1}{W} \sum_{j=w}^W \sqrt{\frac{1}{E} \sum_{e=1}^E \left( \frac{T_{i,e} - T_{j,e}}{T_{i,e} T_{j,e}} \right)^2} \quad \text{dla średnich różnic po oknach względem wybranego okna.}$$

Uwzględniając dyskretny charakter wyników i częstotliwość próbkowania, obliczenia wykonywano wg wzorów

$$(5) \quad \overline{\Delta T_{i,j}} = \frac{1}{rate} \sqrt{\frac{1}{E} \sum_{e=1}^E (Tidx_{i,e} - Tidx_{j,e})^2}$$

i

$$(6) \quad \overline{\overline{\Delta T_i}} = \frac{1}{rate} \frac{1}{W} \sum_{j=B\_H}^{Rect} \sqrt{\frac{1}{E} \sum_{e=1}^E (Tidx_{i,e} - Tidx_{j,e})^2}$$

oraz

$$(7) \quad \overline{\Delta F_{ij}} = rate \sqrt{\frac{1}{E} \sum_{e=1}^E \left( \frac{Tidx_{i,e} - Tidx_{j,e}}{Tidx_{i,e} Tidx_{j,e}} \right)^2}$$

i

$$(8) \quad \overline{\overline{\Delta F_i}} = \frac{rate}{W} \sum_{j=w}^W \sqrt{\frac{1}{E} \sum_{e=1}^E \left( \frac{Tidx_{i,e} - Tidx_{j,e}}{Tidx_{i,e} Tidx_{j,e}} \right)^2}$$

gdzie symbolem  $Tidx$  oznaczono numer składowej cepstralnej wskazujący położenie maksimum związanego z okresem podstawowym  $T0$ , a  $rate$  oznacza częstotliwość próbkowania.



W naszych badaniach liczba rodzajów okien  $W=5$ , liczba przykładów  $E=5400$ , a częstotliwość próbkowania sygnału  $rate = 16 \text{ kHz}$ . Odczyty położenia w cepstrum maksimum związanego z okresem  $T0$ , dzielono więc przez 16, aby uzyskać rezultaty wyrażone w  $ms$ , lub mnożono przez 16000, aby uzyskać wyniki wyrażone w  $Hz$ .

W tabelach 9 i na rysunkach 8. i 9. pokazano odchyłki pomiarów  $T0$  i  $F0$  względem pomiarów uzyskanych z użyciem okien najlepszych (Hamminga, Gaussa), średnio dobrych (Blackmana-Harrisa) i najgorszych (flat-top). Czyli wyniki i wykresy dla

$$(9) \quad \Delta \bar{T}_{i|_{i=HmB\_H|G\_T|F\_T},j} = \sqrt{\frac{1}{E} \sum_{e=1}^E (T_{i,e} - T_{j,e})^2}$$

i

$$(10) \quad \Delta \bar{F}_{i|_{i=HmB\_H|G\_T|F\_T},j} = \sqrt{\frac{1}{E} \sum_{e=1}^E \left( \frac{T_{i,e} - T_{j,e}}{T_{i,e} T_{j,e}} \right)^2}$$

Odniesienie względem okna Hamminga zastosowano spodziewając się, że to okno Hamminga okaże się najlepszym. Okazało się jednak, że najmniej takich różnic jest pomiędzy dedykowanym oknem Gaussa a oknem Hana (tabela 9. Fig. 8., co zobaczymy, gdy za odniesienie przyjmujemy dedykowane pomiarom  $F0$  okno Gaussa.

Analiza wyników wyboru optymalnego parametru  $\sigma$  dla okien Gaussa w rozbiciu wg płci (suplement 1.) ujawnia pewne zróżnicowanie w odniesieniu do okien optymalizowanych dla celów pomiaru  $T0$ , natomiast dla dynamiki płęć osób nie ma znaczenia. Ponieważ jednak poszukujemy okien uniwersalnych, nie wymagających uprzedniej identyfikacji płci, czy wieku osób wybieramy wartość określoną położeniem maksimum ogólnej średniej.

Z badań tych wynika, że kompromisowym oknem stosowalnym zarówno do pomiarów  $T0$  jak i do klasyfikacji dźwięków, będzie okno Blackmana-Harrisa.

Table. 9. Odchyłki odczytów  $T0$  uzyskanych z zastosowaniem poszczególnych okien od odczytów uzyskanych z użyciem okien Hamminga, Blackmana-Harrisa Gaussa dedykowanego dla pomiarów  $F0$  i flat-top

Deviations of the  $T0$  readings obtained with different windows from these obtained with Hamming (a), Blackman-Harris (b), Gauss dedicated to  $F0$  measurements (c) and flat-top (d) windows. See

E:\apl\Spectrum\Windows\_Search\Windows\_Search\_1024\joined\_64\_128\_256\_512\_1024\_4stat.xlsb. Joined differences matrices for all comparisons

a) Hamming windows relative deviations

relative to	T0 deviations [ms]						F0 deviations [Hz]					
	Windows width						Windows width					
Window	64	128	256	512	1024	mean	64	128	256	512	1024	mean
B_H	0,460	1,30	2,98	1,64	1,26	1,53	234	320	333	152	79	224
B_N	0,458	1,29	2,93	1,57	1,26	1,50	233	316	326	145	78	220
F_T	0,520	1,44	3,86	3,93	1,84	2,32	275	377	483	395	130	332
G_S	0,523	1,47	4,08	5,04	5,24	3,27	278	387	545	581	548	468
G_T	0,144	0,339	0,773	0,308	1,31	0,57	73	89	96	29	84	74,1
Hm	0	0	0	0	0	0	0	0	0	0	0	0
Hn	0,390	0,875	1,21	0,580	0,734	0,758	197	210	138	46	50	128
K_B	0,442	1,17	2,36	1,14	1,11	1,24	223	282	265	103	67	188
Rect	0,488	1,13	1,97	1,79	2,91	1,66	245	307	258	145	181	227
mean	0,380	1,003	2,24	1,78	1,74	1,43	195	254	271	177	135	207

## b) Blackman-Harris windows relative deviations

relative to Window	T0 deviations [ms]						F0 deviations [Hz]					
	Windows width						Windows width					
	64	128	256	512	1024	mean	64	128	256	512	1024	mean
B_H	0	0	0	0	0	0	0	0	0	0	0	0
B_N	0,104	0,193	0,606	0,512	0,142	0,311	54	56	75	51	11	49,4
F_T	0,364	0,654	2,34	3,57	1,48	1,68	210	231	362	367	120	258
G_S	0,348	0,618	2,57	4,75	5,10	2,68	201	225	426	564	547	393
G_T	0,461	1,31	3,07	1,61	0,425	1,37	235	323	344	151	37	218
Hm	0,460	1,30	2,98	1,64	1,26	1,53	234	320	333	152	79	224
Hn	0,365	1,01	2,74	1,52	1,06	1,34	194	262	308	146	73	197
K_B	0,246	0,594	1,83	1,16	0,663	0,898	133	164	210	113	53	134
Rect	0,529	1,42	3,39	2,21	2,96	2,10	275	386	402	195	188	289
mean	0,320	0,790	2,17	1,89	1,45	1,32	171	219	273	193	123	196

## c) Gauss optimals windows relative deviations

relative to Window	T0 deviations [ms]						F0 deviations [Hz]					
	Windows width						Windows width					
	64	128	256	512	1024	mean	64	128	256	512	1024	mean
B_H	0,461	1,31	3,07	1,61	0,425	1,37	235	323	344	151	37	218
B_N	0,459	1,30	3,01	1,53	0,449	1,35	233	320	338	143	39	214
F_T	0,522	1,45	3,93	3,92	1,433	2,25	276	379	491	395	117	332
G_S	0,524	1,48	4,15	5,03	5,09	3,25	279	390	552	580	546	469
G_T	0	0	0	0	0	0	0	0	0	0	0	0
Hm	0,144	0,339	0,773	0,308	1,31	0,574	73	89	96	29	84	74,1
Hn	0,396	0,891	1,39	0,483	1,12	0,856	200	217	162	37	78	139
K_B	0,443	1,18	2,47	1,10	0,763	1,19	224	286	280	100	60	190
Rect	0,486	1,12	1,88	1,78	2,95	1,64	245	304	246	147	188	226
mean	0,381	1,008	2,29	1,75	1,50	1,39	196	256	279	176	128	207

## d) Flat-top windows relative deviations

relative to Window	T0 deviations [ms]						F0 deviations [Hz]					
	Windows width						Windows width					
	64	128	256	512	1024	mean	64	128	256	512	1024	mean
B_H	0,364	0,654	2,34	3,57	1,48	1,68	210	231	362	367	120	258
B_N	0,370	0,670	2,43	3,61	1,48	1,71	212	234	369	370	120	261
F_T	0	0	0	0	0	0	0	0	0	0	0	0
G_S	0,296	0,392	0,87	2,95	4,86	1,87	169	151	239	441	536	307
G_T	0,522	1,45	3,93	3,92	1,43	2,25	276	379	491	395	117	332
Hm	0,520	1,44	3,86	3,93	1,84	2,32	275	377	483	395	130	332
Hn	0,434	1,18	3,67	3,88	1,71	2,17	243	331	466	393	126	312
K_B	0,384	0,854	3,01	3,75	1,57	1,91	222	270	413	382	123	282
Rect	0,545	1,52	4,17	4,09	3,22	2,71	287	422	528	395	204	367
mean	0,382	0,908	2,70	3,30	1,96	1,85	211	266	372	349	164	272

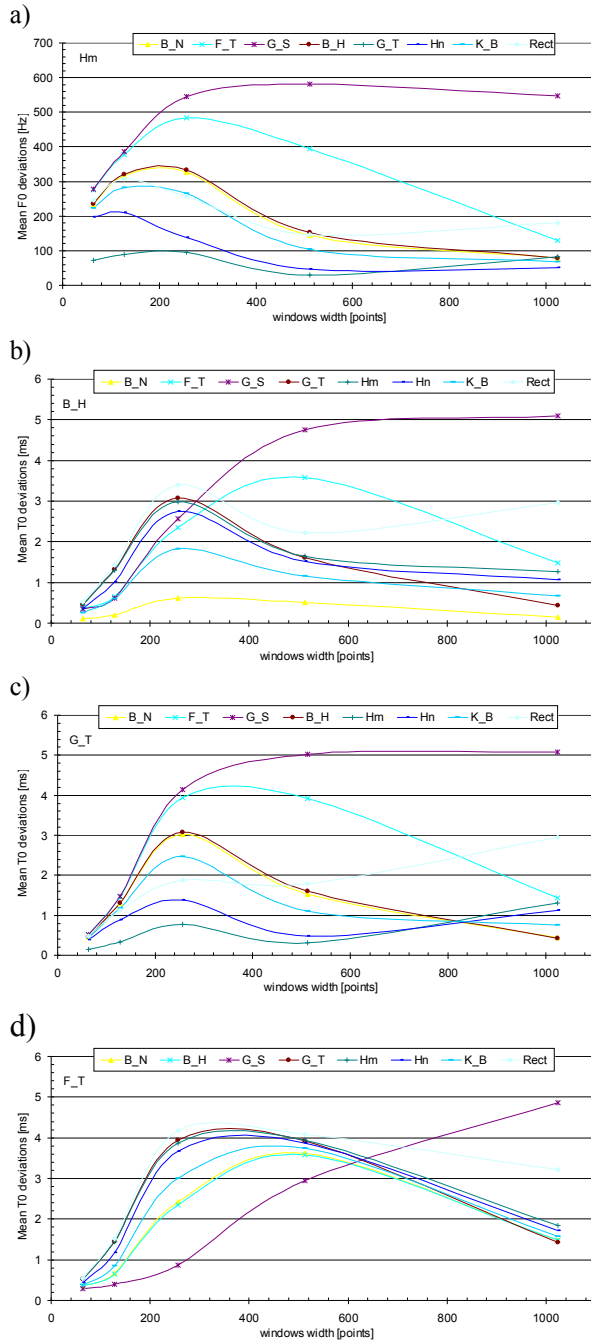


Fig. 8.  $T_0$  readings deviations from the Hamming (a), Blackman-Harris (b) Gauss  $F_0$  dedicated (c) and flat-top (d) windows readings for different windows shapes and width (squared roots of mean of squared deviations) E:\apl\Spectrum\Windows\_Search\Windows\_Search\_64\_128\_256\_512\_1024\joined\_64\_128\_256\_512\_1024\_4stat.xlsb.Joined differences matrices

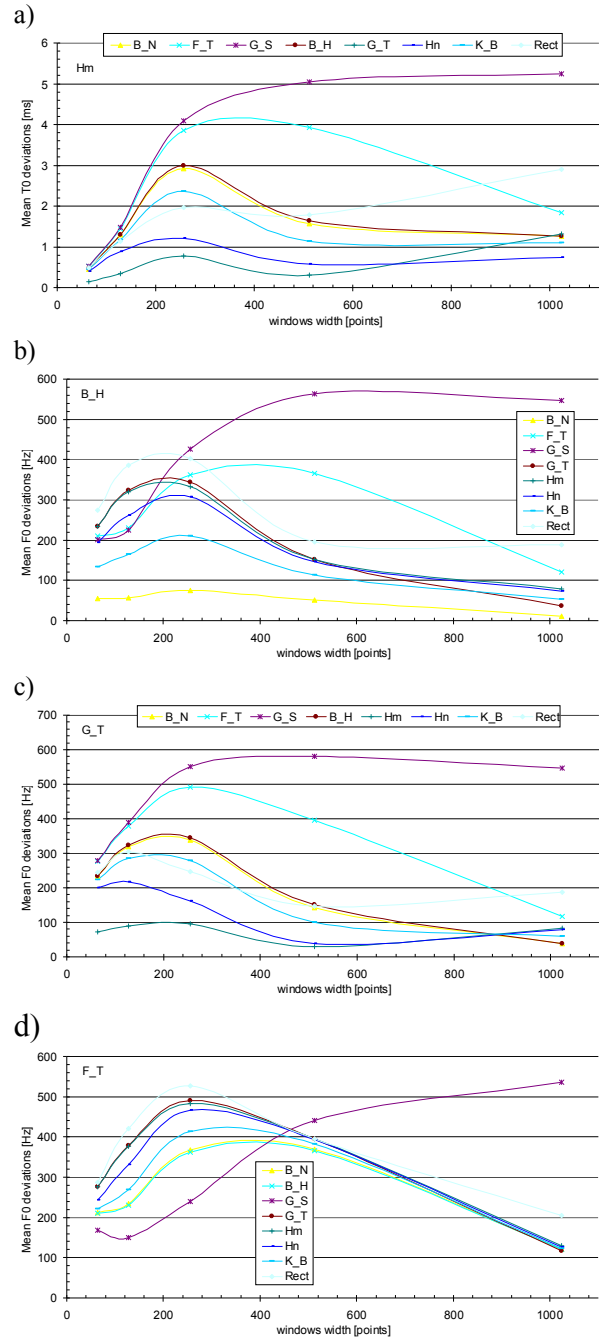
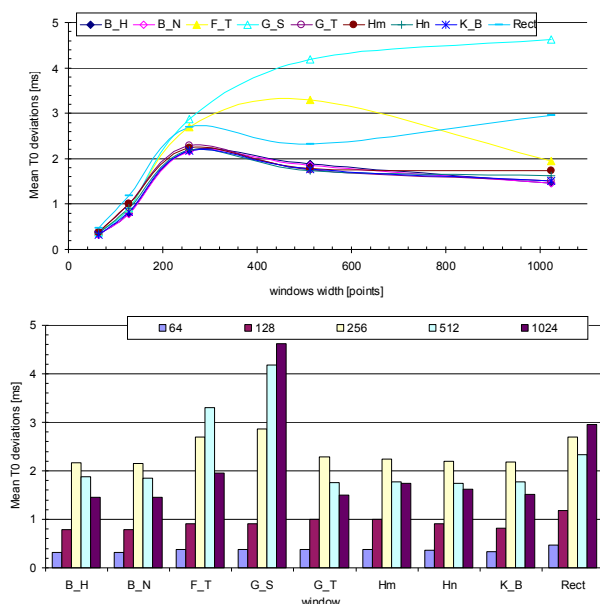
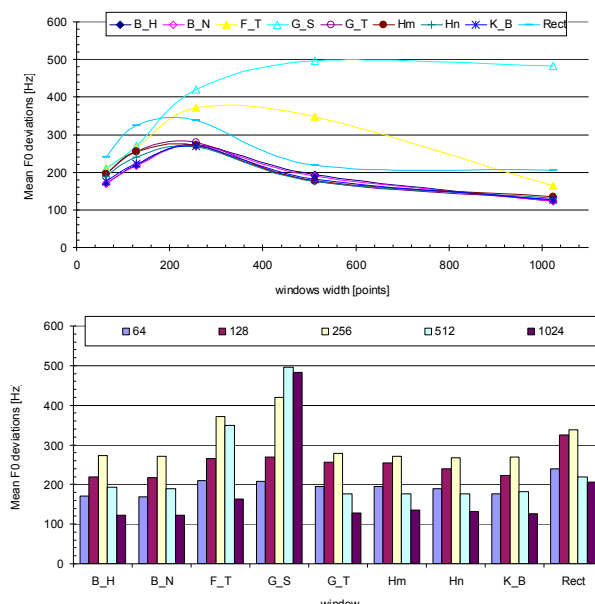


Fig. 9.  $F_0$  readings deviations from the Hamming (a), Blackman-Harris (b), Gauss  $F_0$  dedicated (c) and flat-top (d) windows readings for different windows shapes and width (squared roots of mean of squared deviations) E:\apl\Spectrum\Windows\_Search\Windows\_Search\_64\_128\_256\_512\_1024\joined\_64\_128\_256\_512\_1024\_4stat.xlsb.Joined differences matrices

Table 10. Mean deviations over second windows of  $T0$  and  $F0$  as function of windows widthE:\apl\Spectrum\Windows\_Search\Windows\_Search\_64\_128\_256\_512\_1024\joined\_64\_128\_256\_512\_1024\_4stat.xlsb.Joined differences matrices and  $T0$  differences

T0 means [ms]							F0 means [Hz]						
Windows width							Windows width						
window	64	128	256	512	1024	mean	64	128	256	512	1024	mean	
B_H	0,320	0,790	2,17	1,89	1,45	1,32	171	219	273	193	123	196	
B_N	0,319	0,787	2,15	1,85	1,45	1,31	170	217	271	189	123	194	
F_T	0,382	0,908	2,70	3,30	1,96	1,85	211	266	372	349	164	272	
G_S	0,377	0,912	2,87	4,18	4,62	2,59	209	269	420	496	484	375	
G_T	0,381	1,01	2,29	1,75	1,50	1,39	196	256	279	176	128	207	
Hm	0,380	1,003	2,24	1,78	1,74	1,43	195	254	271	177	135	207	
Hn	0,357	0,908	2,20	1,74	1,62	1,36	189	240	268	176	131	201	
K_B	0,330	0,822	2,18	1,78	1,52	1,32	176	223	270	181	127	195	
Rect	0,463	1,19	2,69	2,33	2,96	1,92	239	325	338	218	205	265	
mean	0,368	0,925	2,39	2,29	2,09	1,61	195	252	307	239	180	235	

Fig. 10. Total mean deviations of  $T0$  as function of windows kind and windows widthE:\apl\Spectrum\Windows\_Search\Windows\_Search\_64\_128\_256\_512\_1024\joined\_64\_128\_256\_512\_1024\_4stat.xlsb.  $T0$  differencesFig. 11. Total mean deviations of  $F0$  as function of windows kind and windows widthE:\apl\Spectrum\Windows\_Search\Windows\_Search\_64\_128\_256\_512\_1024\joined\_64\_128\_256\_512\_1024\_4stat.xlsb.  $T0$  differences

Ten rysunek pokazuje, że najbardziej odstającymi od reszty – jeśli chodzi o rezultaty pomiarów  $T0$ , czy  $F0$  – są okna prostokątne, okna typu flat-top i okna Gaussa optymalizowane dla uwypuklenia dynamiki.

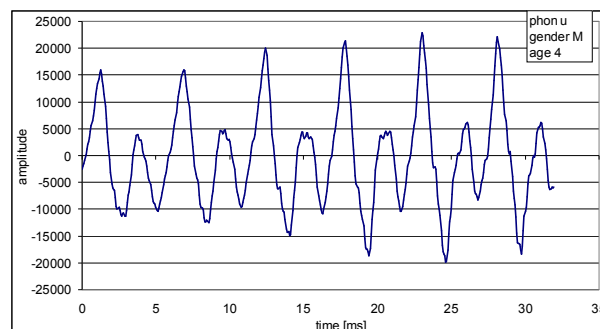
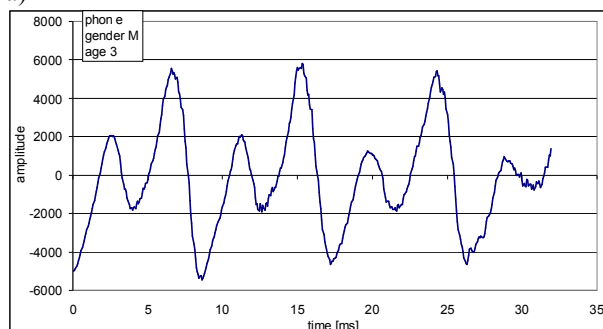
Analiza odchylek sugeruje, że pomiary  $F0$  na podstawie cepstrum są bezużyteczne ze względu na wielkie ich wartości. Jednak pobieżna analiza histogramów odchylek ujawnia, że wartości te są zawyżane tylko przez ok. 2% wyników (okno Hamminga o szerokości 512 pkt.). Jest to efektem błędnego lokowania  $T0$  w wartościach skrajnych cepstrum, w dolnym lub górnym punkcie granicznym obszaru, w którym poszukuje się położenia maksimum związanym z okresem podstawowym, tj. w punkcie liftracji lub w górnym punkcie granicznym zakresu cepstrum. Na rysunku 12 pokazano dwa przykłady takich przypadków. Znalezione je sortując równolegle cały zestaw wyników analiz wg wartości  $T0$  znalezionych przy użyciu okna Hamminga. Wybraliśmy takie przypadki, w których program w poszukiwaniu wartości  $T0$  wchodzi na obszar odpowiedzi impulsowej ( $T0idx=15$ ) i taki, w którym program poszukując

tej wartości wychodzi poza obszar pobudzenia ( $T0idx=256$ )<sub>width=512 pts</sub> i zarazem takie, w których stosowanie innych okien wskazuje inne wartości  $T0$ . Analizy spektralne były wykonane z użyciem okien o szerokości 512 punktów. W metryczkach dołączonych do każdego rysunku zapisano wskazaną wartość  $T0$  oraz wartość cepstrum w tym punkcie.

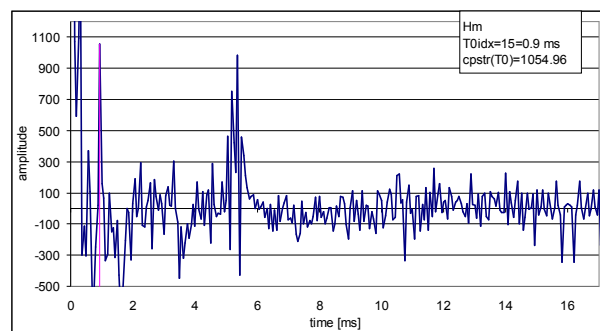
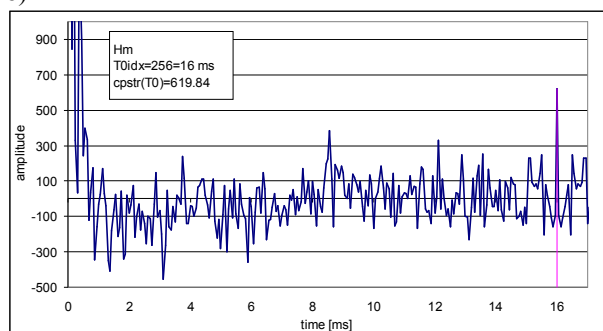
1-th example

2-nd example

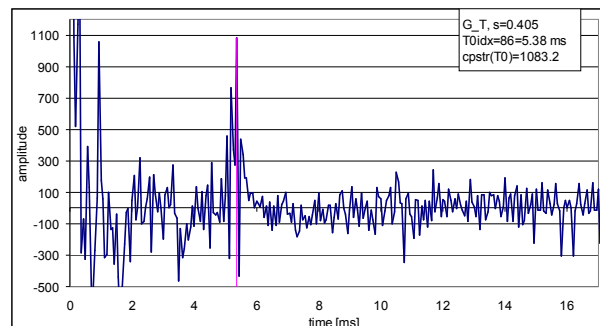
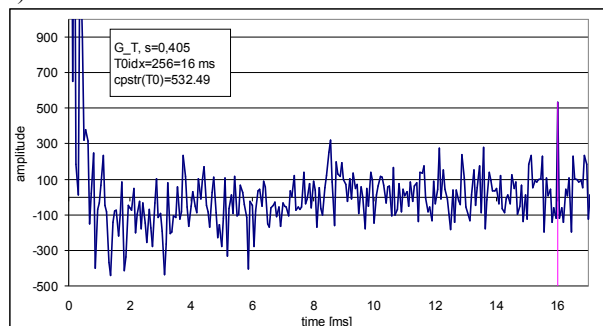
a)



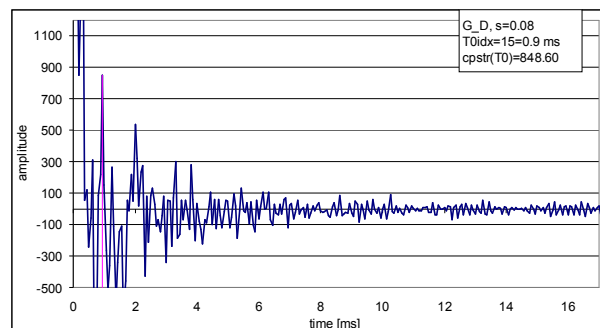
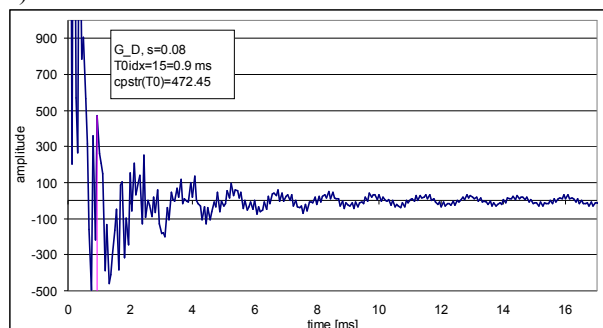
b)



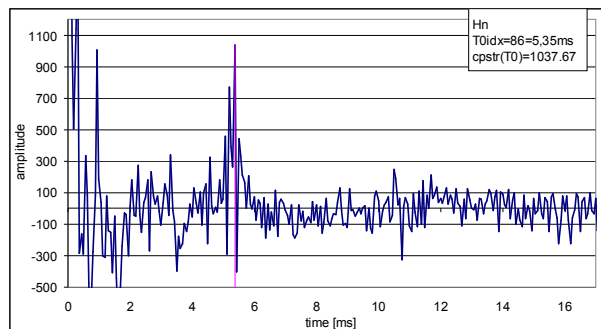
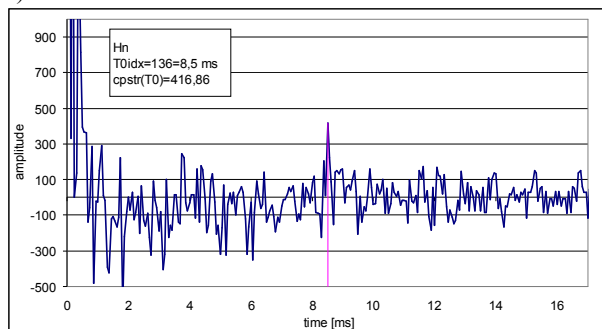
c)



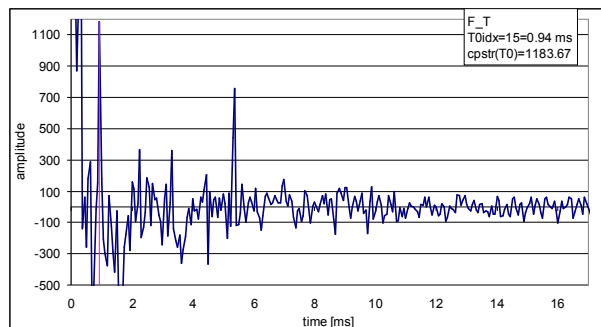
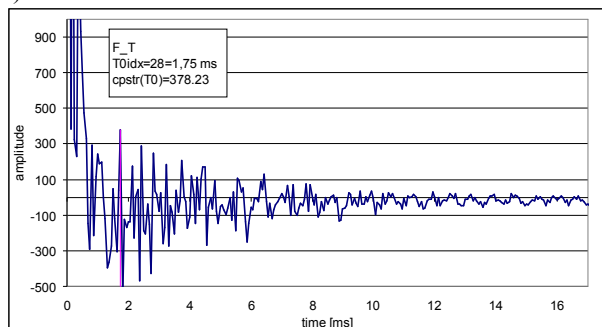
d)



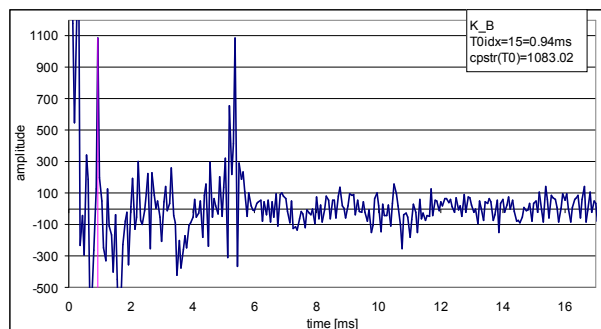
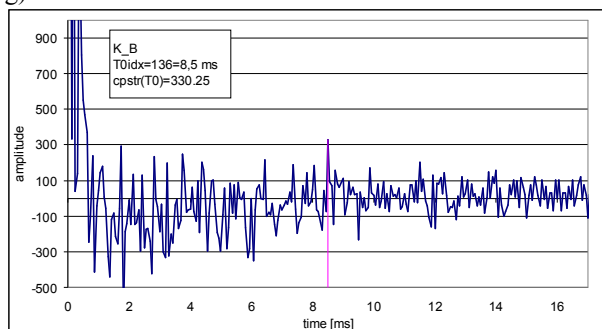
e)



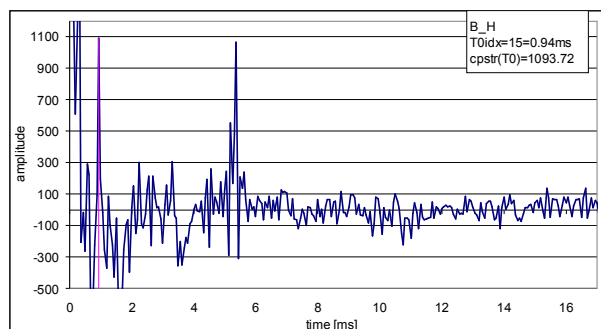
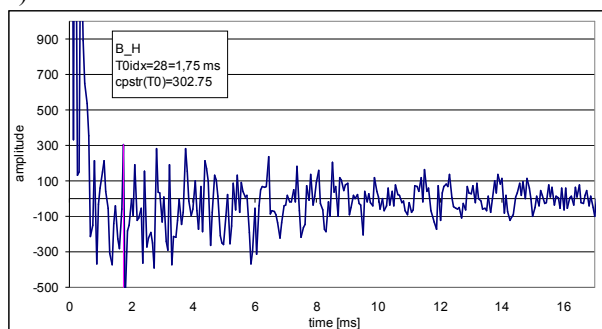
f)



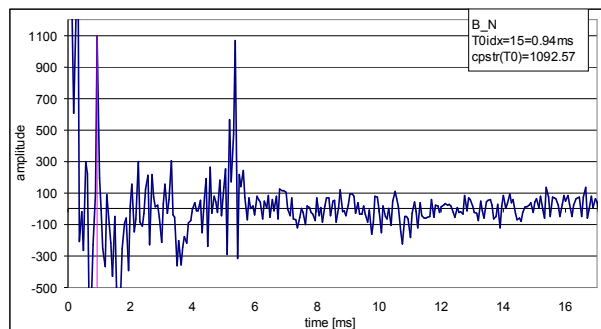
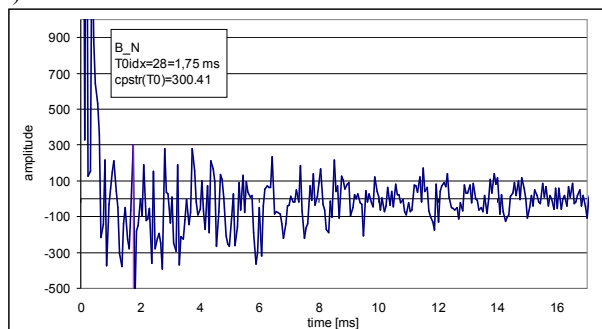
g)



h)



i)



j)

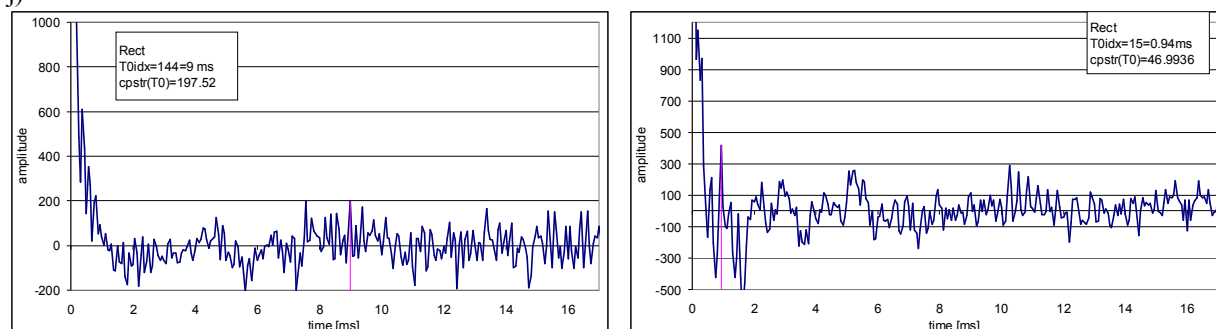


Fig. 12. Two examples of erroneous measurements of  $T_0$ , signal (a) and cepstra obtained using various windows (b-j). Red vertical lines mark  $T_0$  values, windows width=512 points.

1-th example: adresse 234696111, slice number: 2933702, phon: "e", person: "c", gender: "M", age 3

2-nd example: adresse 131452598, slice number: 1643158; phon: "u", person: "T", gender: "M", age: "4".

E:\apl\Spectrum\Windows\_Search\erronous\_T0\_measurements-2-egz.xlsb

Pomiar  $F_0$  za pomocą cepstrum zakłada, że odpowiedź impulsowa toru akustycznego skupia się w niewielu początkowych składowych cepstrum (ok. 12 dla toru głosowego człowieka). Mimo, że u nas próg ten ulokowano w punkcie 15, to i tak zdarza się, że przy zastosowaniu niektórych okien ważących program poszukujący maksimum w obszarze powyżej tego progu wchodzi na obszar odpowiedzi impulsowej znajdując największą wartość cepstrum właśnie w tym punkcie albo bardzo blisko niego. To owocuje oczywiście wielką odchyłką  $T_0$  względem pomiaru uzyskanego za pomocą okna odniesienia. W niektórych przypadkach maksimum to lokowane jest w górnym krańcu cepstrum z takim samym skutkiem. Aby lepiej pokazać ten drugi przypadek, oś odciętych rysunków cepstrów wchodzi na część początkową drugiej połówki tej funkcji (kończy się na 17 ms zamiast na 16, 1-th example).

Przykłady przytoczone na rys. 12 pokazują, że ustalanie wartości  $T_0$  metodą ograniczającą się tylko do poszukiwania największej wartości cepstrum w obszarze pobudzenia jest niewystarczające i wskazują na konieczność stosowania bardziej złożonych procedur ustalania wartości  $T_0$ .

Błędy te skutkowały złym wyborem najlepszej części danej realizacji dźwięku jeśli zdarzały się przy zastosowaniu okna stosowanego w tej procedurze. W naszym przypadku dotyczy to okna Hamminga (wybierano więc nie najlepszą lecz raczej najgorszą część realizacji).

Ze względu na niewielką liczbę takich zdarzeń nie miały one jednak wpływu na wyniki ANOVA.

Skutki tych błędów wymagały dodatkowych analiz. Aby więc bliżej poznać liczbę tego rodzaju błędów w zależności od szerokości i rodzaju okna, wykonano histogramy wartości indeksu  $T_0$  (numerów składowych cepstralnych wskazujących maksymalną wartość cepstrum w obszarze pobudzenia). W suplemencie 5. zamieszczamy tablice liczebności i dystrybuanty wartości  $T_0$ , a poniżej tabele i wykresy odnoszące się do początkowych i końcowych wycinków rozkładów. Za błędne uznajemy pomiary wychodzące poza zakres częstotliwości  $\langle 70, 889 \rangle$

Hz. Dolny kres, 70 Hz, przy częstotliwości próbkowania 16 kHz, jest jednak osiągalny tylko przy użyciu okien o szerokości większej od 512 punktów, więc dla węższych okien, tj. dla okien 64, 128 i 256 punktowych, jako dolną granicę poprawnych pomiarów przyjęto odpowiednio częstotliwości 500, 250 i 128 Hz, wynikające z obliczeń wg wzoru  $f = 2rate / width$ , gdzie  $rate$  oznacza częstotliwość próbkowania, a  $width$  szerokość okna wyrażoną w punktach.



Table 11. Distribution of erroneous measurements of  $F0$ .

E:\apl\Spectrum\Windows\_Search\Windows\_Search\_64\_128\_256\_512\_1024\joined\_64\_128\_256\_512\_1024\_4stat.xlsb.T0 freq. summary

## a) beginning part [%]

w. width (pts)	64	128	256	512	1024	Mean
range index	<=18	<=18	<=18	<=18	<=18	
range (Hz)	>=889	>=889	>=889	>=889	>=889	
B_H	45,8	34,6	17,4	3,9	1,19	<b>20,6</b>
B_N	45,3	34,3	17,1	3,7	1,17	<b>20,3</b>
F_T	51,2	39,7	27,1	15,2	2,24	<b>27,1</b>
G_D	59,9	43,5	33,1	30,0	25,9	<b>38,5</b>
G_T	27,3	21,5	8,48	1,78	1,30	<b>12,1</b>
Hm	27,4	21,8	9,31	1,72	1,31	<b>12,3</b>
Hn	38,2	26,8	10,7	1,85	1,28	<b>15,8</b>
K_B	43,0	31,6	14,5	2,67	1,17	<b>18,6</b>
Rect	24,1	14,1	3,54	1,96	3,91	<b>9,52</b>
<b>Mean</b>	<b>40,3</b>	<b>29,8</b>	<b>15,7</b>	<b>6,97</b>	<b>4,38</b>	<b>19,4</b>

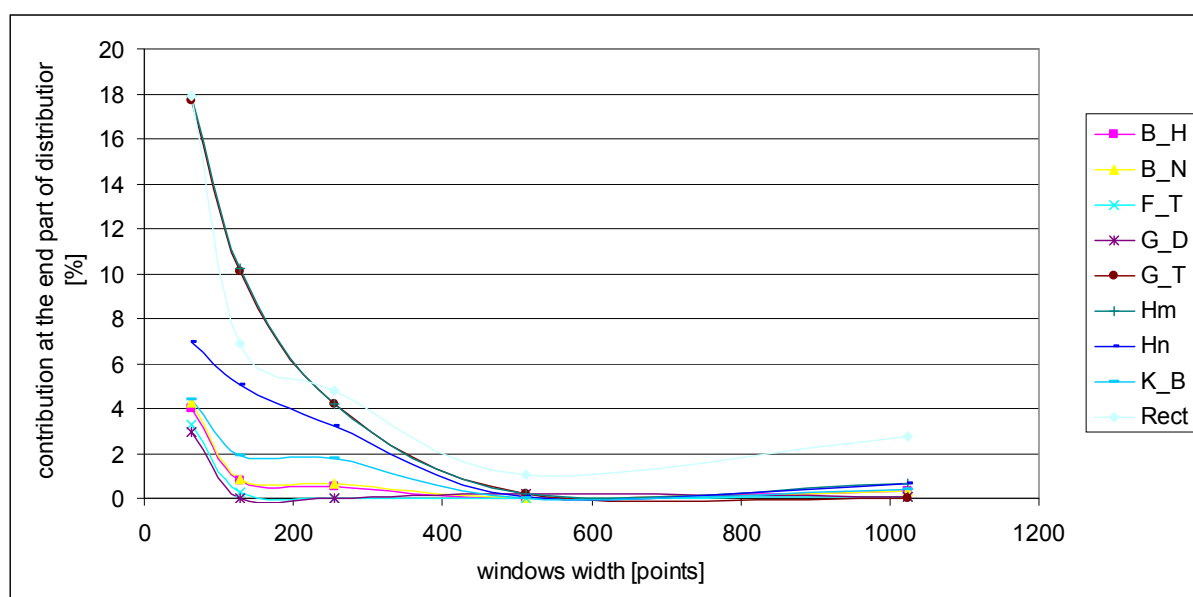
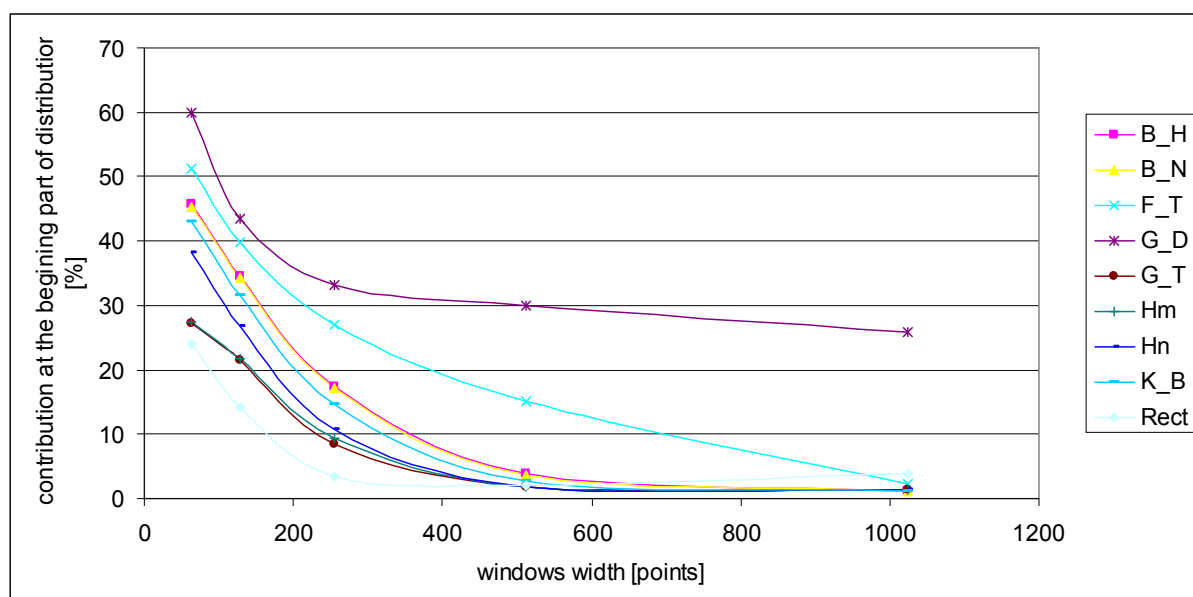
## b) end part [%]

w. width (pts)	64	128	256	512	1024	Mean
range index	32	64	128	>=228	>=228	
range (Hz)	500	250	125	<=70	<=70	
B_H	4,00	0,76	0,519	0,019	0,315	<b>1,12</b>
B_N	4,26	0,83	0,630	0,019	0,315	<b>1,21</b>
F_T	3,28	0,28	0	0,019	0,074	<b>0,73</b>
G_D	2,94	0	0	0,167	0,074	<b>0,64</b>
G_T	17,7	10,1	4,20	0,167	0,003	<b>6,44</b>
Hm	17,9	10,2	4,17	0,185	0,685	<b>6,63</b>
Hn	6,93	5,02	3,20	0,037	0,667	<b>3,170</b>
K_B	4,41	1,87	1,76	0,019	0,426	<b>1,70</b>
Rect	17,9	6,9	4,80	1,037	2,759	<b>6,68</b>
<b>Mean</b>	<b>8,81</b>	<b>4,00</b>	<b>2,14</b>	<b>0,185</b>	<b>0,591</b>	<b>3,15</b>

## c) summ of beginning and end part of distributions [%]

w. width (pts)	64	128	256	512	1024	Mean
sum	sum	sum	sum	sum	sum	
B_H	49,8	35,4	17,9	3,93	1,50	<b>21,7</b>
B_N	49,6	35,1	17,7	3,69	1,48	<b>21,5</b>
F_T	54,5	40,0	27,1	15,2	2,31	<b>27,8</b>
G_D	62,9	43,5	33,1	30,1	26,0	<b>39,1</b>
G_T	45,0	31,6	12,7	1,94	1,30	<b>18,5</b>
Hm	45,3	32,1	13,5	1,91	2,00	<b>18,9</b>
Hn	45,1	31,8	13,9	1,89	1,94	<b>18,9</b>
K_B	47,4	33,4	16,3	2,69	1,59	<b>20,3</b>

Rect	42,0	21,0	8,33	3,00	6,67	16,2
Mean	49,1	33,8	17,8	7,16	4,97	22,6



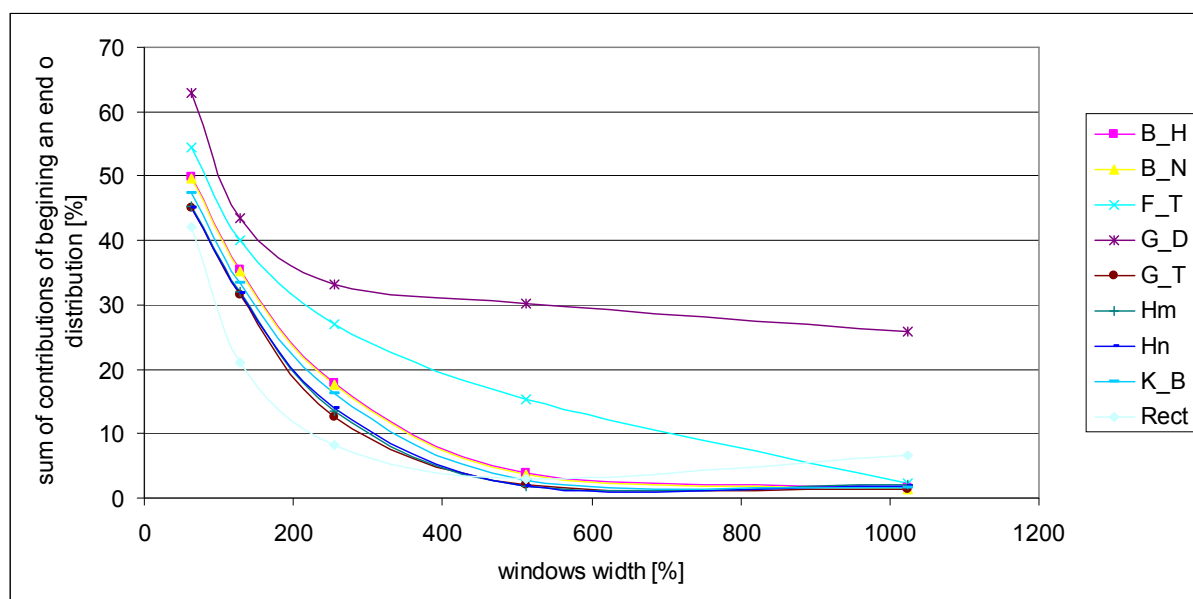


Fig. 13. Graphs of contributions of erroneous measurements of  $F_0$

E:\apl\Spectrum\Windows\_Search\Windows\_Search\_64\_128\_256\_512\_1024\joined\_64\_128\_256\_512\_1024\_4stat.xlsb. T0 freq. summary

**Uwagi.** Stosując okna o różnych szerokościach badamy **różne** wycinki każdej danej realizacji dźwięku nie tylko dlatego, że są one coraz szersze, ale i dlatego, że może być różne ich przesunięcie względem środka danej realizacji dźwięku. Wycinki te, wyizolowane oknami o różnych szerokościach, nie koniecznie będą ułożone koncentrycznie. Jednak naszym zadaniem jest porównanie skuteczności okien w zależności od kształtu, a nie w zależności od analizowanych wycinków; tu dążyliśmy jednak do wyboru do analiz wycinków najlepszych ze względu na wartość cepstrum w punkcie  $T_0$ .

## References

- [1] R. B. Randall, 1987. Frequency analysis. Brüel&Kjær, K. Larsen&Søn, Denmark
- [2] A. Stanis, 2007. Przystępny kurs statystyki z zastosowaniem Statistica Pl na przykładach z medycyny. Statsoft Polska, Kraków, Poland
- [3] S. Grocholewski, 1997. Baza nagrań sygnałów mowy CORPORA. Instytut Informatyki, Politechnika Poznańska, Poznań, Poland