POLITECHNIKA GDAŃSKA
Wydział Elektroniki, Telekomunikacji i Informatyki

Romuald Mazurek

Przestrzenno-czasowe rozkłady pola akustycznego zespołów źródeł szerokopasmowych i ich wpływ na zniekształcenia sygnałów mowy

Rozprawa doktorska

Promotor:
dr hab. inż. Henryk Lasota
Wydział Elektroniki, Telekomunikacji i Informatyki
Politechnika Gdańska

Gdańsk, 2014

Ewie, Annie i Jakubowi

Spis treści

Spis oznaczeń

a	– wektor współczynników predykcji
$a(\xi)$	– funkcja aperturowa dla pobudzenia równomiernego
$\{a_1, \ldots, a_p\}$	– współczynniki predykcji LPC
a_{IN}	– wektor współczynników predykcji sygnału oryginalnego
a_{OUT}	– wektor współczynników predykcji sygnału zniekształconego
A	– względna średnica apertury
A_N	– poziom *N*-tego formantu
$A(z)$	– filtr odwrotny do filtra sygnału mowy (filtr inwersyjny, filtr „wybielający”)
$c_m^c(i)$	– współczynniki cepstralne sygnału oryginalnego
$c_m^d(i)$	– współczynniki cepstralne sygnału zniekształconego
d	– rozstaw źródeł
$d(a,b)$	– ogólna miara odległości między wektorami
$d_{LLR}(\mathbf{a}_d, \mathbf{a}_c)$	– miara odległości LLR
$d_{LLRS}(\mathbf{a}_d, \mathbf{a}_c)$	– symetryczna miara odległości LLR
$d_{IS}(\mathbf{a}_d, \mathbf{a}_c)$	– miara odległości IS
$d_{LPCCD}{}^2(i)$	– odległość cepstralna LPC CD
$D(\vartheta, \omega)$	– znormalizowana funkcja rozkładu kątowo-częstotliwościowego
$D_\vartheta(\omega)$	– funkcja przenoszenia dla ustalonego kierunku
$D_\omega(\vartheta)$	– charakterystyka kierunkowa dla ustalonej częstotliwości
$\delta[n]$	– delta Kroneckera
σ_c^2	– wariancja błędu predykcji sygnału wzorcowego
σ_d^2	– wariancja błędu predykcji sygnału zakłóconego
$e[n]$	– błąd predykcji
$\{e[n]\}$	– sygnał resztkowy
E	– minimalny błąd średniokwadratowy
$E(z)$	– *z*-transformata sygnału resztkowego,
f	– częstotliwość
f_{mel}	– częstotliwość w skali melowej
F_0	– ton podstawowy (ton krtaniowy)
F1	– częstotliwość pierwszego formantu
F2	– częstotliwość drugiego formantu
F_{mod}	– częstotliwość modulująca

G – parametr wzmocnienia w kodeku LPC

$G(X)$ – wielomian okresowego ciągu bitów

G(z) – transmitancja filtru resztkowego

$\gamma(m)$ – okno czasowe obserwacji

$h(\vec{x},t)$ – odpowiedź impulsowa w punkcie odsłuchu

$h(r, t)$ – funkcja odpowiedzi impulsowej o symetrii sferycznej

$h_N(\vec{x},t)$ – całkowita odpowiedź impulsowa w systemie *N*-źródłowym

$h_i(r_i,t)$ – odpowiedź impulsowa dla *i*-tego źródła

$h_g(t)$ – odpowiedź impulsowa głośnika

$h_{Ng}(\vec{x},t)$ – całkowita odpowiedź impulsowa w systemie *N*-głośnikowym

$h(r_0, \vartheta, t)$ – odpowiedź impulsowa w strefie dalekiej pola

$H(\vec{x},\omega)$ – uogólniona funkcja przenoszenia układu apertura promieniująca – obserwator

$H_{r_0}(\vartheta,\omega)$ – uogólniona funkcja przenoszenia jako funkcja kąta i pulsacji

$H_{Ng}(\vec{x},\omega)$ – funkcja przenoszenia w systemie *N*-głośnikowym

$H(z)$ – transmitancja filtru odwrotnego

$I(\vec{x})$ – natężenie dźwięku

K – liczba formantów, jaka jest brana pod uwagę w modelu

L – wartość stosunku sygnału do szumu

L_F – numer prążka odpowiadający częstotliwości modulującej

$\mathrm{M}(\vec{x})$ – punkt obserwacji

$MFCC_n$ – *n*-ty współczynnik mel-cepstralny

$MTF(F_{\mathrm{mod}})$ – wskaźnik funkcji przeniesienia modulacji w funkcji częstotliwości modulującej

MTF_j – wskaźnik funkcji przeniesienia modulacji dla *j*-tego podpasma oktawowego

N – liczba źródeł

p – rząd predykcji LPC

$p(L, STR, s)$ – prawdopodobieństwo prawidłowego rozpoznania

r – odległość punktu odsłuchu od źródła

r – wektor współczynników autokorelacji

r_F – granica strefy dalekiej

r_i – odległość *i*-tego głośnika od punktu odsłuchu

r_k – *k*-ty współczynnik autokorelacji

R – macierz autokorelacji

$R_{xy}[n]$ – funkcja korelacji skrośnej sygnałów x[n] oraz y[n]

$R_{xx}[n]$ – funkcja autokorelacji sygnału pomiarowego

$s(t)$ – sygnał pobudzający

S – nachylenie wykresu zrozumiałości

S(z) – *z*-transformata sygnału oryginalnego

$STFT(n,k)$ – krótkoczasowa dyskretna transformata Fouriera

S_i – uśredniona wartość estymaty widma gęstości mocy przy użyciu *i*-tego filtru

t – czas bieżący

t_i – czas przelotu od *i*-tego źródła do punktu obserwacji

T – okres trwania sygnału

$\{w_i\}$ – funkcja okna obejmującego *N* próbek sygnału

W(j, m) – waga *j*-tego podpasma w *m*-tym segmencie

W – średnia wyrazistość logatomowa

W_i – wskaźnik wyrazistości w *i*-tym pasmie elementarnym

$W_{n,l}$ – liczba logatomów rozpoznanych poprawnie przez *n*-tego słuchacza

ω_0 – pulsacja, dla której funkcja przenoszenia w kierunku ϑ ma pierwsze zero

ϑ_0 – kierunek, w którym charakterystyka kierunkowa ma pierwsze zero

x[n] – próbka sygnału wzorcowego

$\{\hat{x}[n]\}$ – sygnał estymowany

$x'[n]$ – sygnał przefiltrowany nierekursywnym filtrem FIR (preemfaza)

$X^{0\ldots N}$ – zawartość komórki rejestru

$y(\vec{x},t)$ – sygnał w punkcie odsłuchu

Wykaz akronimów

ACR	(Absolute Category Rating) - metoda bezwzględnej oceny jakości mowy
AI	(Articulation Index) – wskaźnik wyrazistości mowy
AR	(AutoRegresive) – model autoregresyjny
ARM	Automatyczne Rozpoznawanie Mowy
AWGN	(Additive White Gaussian Noise) - biały szum Gaussowski
CCR	(Comparision Category Rating) - metoda porównawcza oceny jakości mowy
CD	(Cepstrum Distance) – odległość cepstralna
CIS	(Common Intelligibility Scale) – wspólna skala zrozumiałości
CMOS	(Comparison Mean Opinion Score) - porównawcza uśredniona opinia słuchaczy
CORPORA	baza referencyjne („korpus”) mowy polskiej
DCR	(Degradation Category Rating) - metoda oceny stopnia degradacji jakości mowy
DMOS	(Degradation Mean Opinion Score) – degradacyjna uśredniona opinia słuchaczy
DFT	(Discreet Fourier Transform) – dyskretna transformata Fouriera
DRT	(Diagnostic Rhyme Test) - diagnostyczny test rymowy
DSO	Dźwiękowe Systemy Ostrzegawcze
IP	(Internet Protocol) – protokół transmisji danych w Internecie
IS	(Itakura-Saito) – miara odległości Itakura-Saito
ISP	(Inverse Sine Parameters) - parametry odwrotnej funkcji sinus
FFT	(Fast Fourier Transform) – szybki algorytm transformaty Fouriera
FHT	(Fast Hadamard Transform) – szybki algorytm transformaty Hadamarda
FIR	(Finite Impulse Response) – filtr o skończonej odpowiedzi impulsowej
fwSNRseg	(Frequency Weighted SNR) – ważony częstotliwościowo stosunek do szumu
LAR	(Log Area Ratio Parameters) - współczynniki logarytmicznego stosunku przekrojów tuby akustycznej
LFS	(Line Spectral Frequencies) - metoda częstotliwości widma liniowego
LFSR	(Linear Feedback Shift Register) - liniowy rejestr przesuwny ze sprzężeniem zwrotnym
LLR	(Log-Likelihood Ratio) – logarytmiczny wskaźnik wiarygodności
LPC	(Linear Predictive Coding) – liniowe kodowanie predykcyjne
LPC CD	(LPC Cepstrum Distance) – odległość cepstralna przy kodowaniu LPC
LPS	(Line Spectrum Pair) - metoda par widma liniowego
LR	(Likelihood Ratio) – wskaźnik wiarygodności
LTI	(Linear Time Invariant) – układ liniowy niezmienny w czasie
melCD	(MFCC Distance) – odległość dla współczynników mel-cepstralnych
MFCC	(Mel-Frequency Cepstral Coefficients) – współczynniki mel-cepstralne

MLS	(Maximum Length Sequence) – ciąg maksymalnej długości
MOS	(Mean Opinion Score) - uśredniona opinia słuchaczy
MRT	(Modified Rhyme Test) – zmodyfikowany diagnostyczny test rymowy
MTF	(Modulation Transfer Function) – funkcja przeniesienia modulacji
PAMS	(Perceptual Analysis Measurement System) – system pomiaru z analizą perceptualną
PAS	(Public Address Systems) - systemy dźwiękowe instalowane w miejscach publicznych
PEAQ	(Perceptual Evaluation of Audio Quality) – badanie jakości dźwiękowych sygnałów szerokopasmowych
PESQ	(Perceptual Evaluation of Speech Quality) – badanie jakości sygnałów mowy
PIR	(Periodic Impulse Response) – okresowa odpowiedź impulsowa
PLP	(Perceptual Linear Prediction) - predykcja liniowa uwzględniająca podejście perceptualne
PRN	(PseudoRandom Noise) - pseudolosowy przebieg szumowy
PSD	(Power Spectral Density) – gęstość widmowa mocy
PSQM	(Psycho-Acoustic Speech Quality Measure) – metoda psycho-akustycznego pomiaru jakości mowy
PSQM+	(Psycho-Acoustic Speech Quality Measure Plus) – metoda psycho-akustycznego pomiaru jakości mowy w otoczeniu sieciowym
PSTN	(Public Switched Telephone Network) - publiczna komutowana sieć telefoniczna
RASTA	(RelAtive SpecTraA) - metoda parametryzacji wykorzystująca widmo względne
RASTI	(RApid Speech Transmission Index) – szybki wskaźnik zrozumiałości mowy
SFM	(Spectral Flatness Measure) – miara płaskości widma
SII	(Speech Inteligibility Index) - wskaźnik zrozumiałości mowy
SRT	(Speech Reception Threshold) – próg percepcji mowy
STFT	(Short-Time Fourier Transform) – krótkoczasowa transformata Fouriera
STI	(Speech Transmission Index) – wskaźnik transmisji mowy
STIPA	(Speech Transmission Index for Public Address Systems) - wskaźnik transmisji mowy w systemach publicznych
SNR	(Signal to Noise Ratio) – stosunek sygnału do szumu
SNRseg	(Time-Domain Segmental SNR) – segmentowy stosunek sygnału do szumu
SPL	(Sound Pressure Level) – poziom ciśnienia akustycznego
SQ	(Speech Quality) – jakość sygnału mowy
WAV	(Wave Form Audio Format) - format plików dźwiękowych
VoIP	(Voice over Internet Protocol) – protokół przesyłania sygnału mowy za pomocą łączy internetowych IP
%Alcons	(Articulation Loss of Consonants) - współczynnik utraty spółgłosek

1. WPROWADZENIE

Podstawowym fizycznym mechanizmem formowania pola akustycznego w przestrzeni jest zachodząca w każdym jej punkcie liniowa superpozycja chwilowych wartości ciśnienia. Gdy pole akustyczne jest kreowane przez źródła pobudzane niemal identycznym sygnałem, o niewielkich różnicach wynikających z indywidualnych cech źródeł, dominującym zjawiskiem jest interferencja, której wpływ na sygnały szerokopasmowe nie jest dobrze poznany [1] [2] [3].

Uzasadnieniem stosowania układów wielogłośnikowych jest tzw. kierunkowość, którą powinny charakteryzować się układy złożone z wielu źródeł, co w założeniu wielu projektantów powinno ograniczać wypromieniowywanie energii do obszaru, w którym nie ma odbiorców [4] [5]. W praktyce założenie o kierunkowości tego typu apertur nie jest zawsze prawdziwe. W obszarze bliższym źródłu, noszącym nazwę strefy bliskiej, rozkłady pola mają charakter na tyle złożony, że nie można tam określić funkcji rozkładu o podobnie jednoznacznych właściwościach, jak charakterystyka kierunkowa. Stosowanie wieloźródłowych układów akustycznych, szczególnie rozłożonych wzdłuż linii, uzasadnia się również względami estetycznymi, gdyż wkomponowują się one w architekturę pomieszczenia, np. poprzez umieszczanie ich na kolumnach podporowych.

1.1. Istota problemu

W akustycznych systemach wieloźródłowych pobudzanych wspólnym szerokopasmowym sygnałem dźwiękowym występuje zjawisko interferencji powodujące znaczne zniekształcenia liniowe. Badanie układów akustycznych złożonych z zespołu wielu źródeł szerokopasmowych, traktowanych w kategoriach systemów liniowych LTI (*ang. Linear Time Invariant*) i zastosowanie metody odpowiedzi impulsowych, pozwala określić charakter zniekształceń powstających w polu tego typu układów [6].

Zjawisko to (*interferencja szerokopasmowa*) będące ekstrapolacją interferencji sygnałów harmonicznych na sygnały szerokopasmowe jest szczególnie silne w obszarach nagłaśnianych przez większą liczbę źródeł znajdujących się w porównywalnych odległościach. Efektem są zniekształcenia liniowe powodujące znaczne, w stosunku do oryginału, różnice postaci czasowej i widmowej sygnałów docierających do poszczególnych miejsc w polu odsłuchu i, w konsekwencji, pogorszenie zrozumiałości przekazu. Zjawiska tego nie można pomijać przy projektowaniu oraz badaniu jakości systemów nagłaśniających.

Chociaż zagadnienie formowania się pola akustycznego [7] w nagłaśnianej przestrzeni jest analogiczne do formowania pola wieloelementowych anten w radiowych lub hydroakustycznych (ultradźwiękowych) systemach komunikacji bezprzewodowej, to z punktu widzenia dostępnych narzędzi analizy jest jakościowo inne. Anteny promieniują bowiem sygnały wąskopasmowe, które można traktować jako czysto harmoniczne, natomiast sygnał mowy jest z natury szerokopasmowy, jego przebieg czasowy jest złożony, a widmo częstotliwościowe pokrywa w praktyce około dwie

i pół dekady (od 50 Hz do 15 kHz). Do anten szerokopasmowych nie można więc stosować klasycznego opisu w rodzaju charakterystyk kierunkowych, które to z definicji dotyczą pola harmonicznego [4].

Dodatkowo, w analizowanym problemie konfiguracja geometryczna elementów promieniujących nie jest regularna, a punkt obserwacji (słuchacz) znajduje się w obszarze pomiędzy tymi elementami. Są to kolejne powody, dla których nie jest możliwe operowanie funkcjami typu charakterystyki kierunkowe – te, bowiem, dotyczą typowych konfiguracji i obowiązują w obszarze dostatecznie odległym od anteny, w stosunku do względnego rozmiaru jej apertury (tzw. strefa daleka).

Typowe systemy publicznego przekazu słownego PAS (*ang. Public Address Systems*) oraz dźwiękowe systemy ostrzegawcze DSO [8] wykorzystują zwielokrotnione źródła w postaci większej liczby głośników rozmieszczonych w możliwie równomierny sposób w obszarze odsłuchu, najczęściej na powierzchniach ograniczających nagłaśniane pomieszczenie (ściany, sufit) lub na elementach konstrukcyjnych (filary, pilastry).

Uwaga projektantów systemów nagłośnieniowych zogniskowana jest na tzw. akustyce pomieszczenia, w którym realizowany jest przekaz słowny. Zjawiska pogłosowe związane z geometrią sali audytoryjnej, wynikające z wielokrotnych odbić fali dźwiękowej uznawane są za dominujące [9] [10]. Tymczasem zjawisko interferencji towarzyszące generowaniu pola dźwiękowego przez źródła pobudzane identycznym sygnałem powoduje znaczne zniekształcenia liniowe. Zrozumiałość przekazu słownego może ulec pogorszeniu, zależnemu w silnym stopniu od liczby źródeł dźwięku oraz od ich rozmieszczenia względem miejsca odsłuchu.

Stosowane powszechnie ustandaryzowane metody pomiarów i oceny zrozumiałości mowy w pomieszczeniach np. STI, RASTI, stosunkowo dobrze odzwierciedlają wpływ zjawisk pogłosowych na zrozumiałość [10] [11] [12], są one jednak, jak to pokazano w niniejszej pracy, niewrażliwe na wpływ zjawiska superpozycji sygnałów dochodzących do słuchacza z wielu źródeł, a które to zjawisko wpływa istotnie na jakość odbieranych sygnałów dźwiękowych. Wpływ ten uwidacznia się szczególnie w warunkach wysokiego poziom szumu tła. Właściwe opisanie rodzaju i charakteru tego zjawiska, określenie mierzalnej cechy zniekształceń oraz wykazanie ich wpływu na jakość pola akustycznego umożliwi oszacowanie stopnia degradacji i zniekształcenia transmitowanego w polu sygnału akustycznego, a w szczególności przekazu mowy.

Projektowanie systemów nagłośnienia PAS i DSO wymaga opracowania nowego, kompleksowego podejścia do problemu. Własności odsłuchowe, zwyczajowo określane w sposób ogólny, przez podanie parametrów zbiorczych odnoszących się do całego pomieszczenia, należałoby dodatkowo charakteryzować parametrami lokalnymi, uwzględniającymi zniekształcenia specyficzne dla konkretnej konfiguracji głośników i wybranej lokalizacji słuchacza. Obiektywna

ocena wpływu zniekształceń liniowych na pogorszenie zrozumiałości mowy powinna uwzględniać parametry przekazu mowy na poziomie fonemów.

1.2. Tezy i cele rozprawy

Badanie wpływu rozmieszczenia i liczby źródeł dźwięku na zrozumiałość sygnałów mowy jest kluczowe dla projektowania systemów nagłośnienia. Efekt interferencji szerokopasmowej oraz jej wpływ na zrozumiałość przekazywanych komunikatów słownych są ważnym elementem uzupełniającym kompleksowy opis zjawisk akustycznych związanych z nagłaśnianiem wraz z takimi czynnikami jak wpływ pomieszczenia (pogłos, tłumienie) czy jakość przetworników. Analiza zagadnienia przedstawiona w pracy może w konsekwencji dalszych badań doprowadzić do opracowania zaleceń uzupełniających już istniejące.

Celem głównym pracy jest potwierdzenie prawdziwości postawionych tez, które brzmią następująco:

1. **Superpozycja sygnałów pochodzących ze współbieżnych źródeł rozmieszczonych w różnych odległościach od punktu odsłuchu wywołuje efekt zniekształcenia parametrów czasowo-częstotliwościowych sygnałów szerokopasmowych**
2. **Wskazane jest zastosowanie obiektywnej, powtarzalnej miary degradującego wpływu szerokopasmowej interferencji liniowej na zrozumiałość sygnałów mowy.**

Aby wykazać słuszność powyższych tez, postawiono następujące cele pomocnicze:

- Sformułowanie problemu szerokopasmowych układów akustycznych w kategoriach systemów liniowych LTI (ang. *Linear Time Invariant*).
- Przeprowadzenie analizy czasowo-częstotliwościowej układów akustycznych przy wykorzystaniu metody odpowiedzi impulsowej.
- Identyfikacja zniekształceń powstających na skutek rozmieszczenia źródeł akustycznych w przestrzeni odsłuchowej.
- Przegląd stosowanych metod parametryzacji i badania jakości sygnałów mowy oraz dobór wskaźników adekwatnych dla pomiaru zniekształceń wprowadzanych przez interferencję szerokopasmową.
- Skonstruowanie modelu symulującego transmisję sygnałów szerokopasmowych w złożonych układach przestrzennych rozkładów źródeł oraz weryfikującego wartości obiektywnych parametrów sygnałów w oparciu o analizę predykcyjną i mel-cepstralną.
- Przeprowadzenie badań pola akustycznego typowych układów przestrzennego rozkładu źródeł akustycznych z wykorzystaniem miar odległości wskaźników wiarygodności: LLR (Log-Likelihood Ratio), IS (Itakura-Saito), CD (Cepstrum Distance), melCD (MFCC Distance).

- Wykazanie korelacji pomiędzy przestrzennym rozkładem źródeł a zniekształceniami parametrów sygnału mowy.
- Dokonanie oceny zniekształceń oraz ich wpływu na wartości wskaźników odległości, mających cechę testowania istotność wpływu interferencji szerokopasmowej na degradację parametrów sygnału mowy.

Dla weryfikacji koncepcji przeprowadzono testy pomiarowe pola akustycznego rzeczywistych układów źródeł akustycznych.

Dla realizacji tak postawionych celów niezbędne było przyjęcie następujących założeń:

- Układy akustyczne, złożone z zespołów źródeł szerokopasmowych, mogą być analizowane w kategoriach systemów liniowych LTI.
- Wykorzystanie do analizy metody odpowiedzi impulsowej pozwala określić charakter zniekształceń powstających w polu układów tego typu.
- Wpływ przestrzennego rozkładu źródeł na jakość transmitowanego sygnału należy rozważać uwzględniając przede wszystkim zniekształcenia parametrów formantowych przekazu głosowego.

Badania potwierdzające zasadność postawionych tez przeprowadzono przy założeniu braku wpływu innych czynników powodujących zniekształcenia sygnału, takich jak:

- nierównomierność charakterystyk częstotliwościowych przetworników elektroakustycznych,
- kierunkowość pojedynczego źródła akustycznego,
- odbicia i rewerberacje w pomieszczeniach nagłaśnianych,
- zniekształcenia nieliniowe sygnału powstające w urządzeniach akustycznych (np. wzmacniaczach akustycznych, głośnikach),
- absorpcja akustyczna szczególnie istotna dla dużych częstotliwości.

Tradycyjnie uwzględniany wpływ tych czynników na badanie jakości przekazu mowy w nagłaśnianych pomieszczeniach jest szeroko opisany w literaturze tematu. Powyższe elementy pominięto jako nie mające bezpośredniego związku z badanym problemem, a pozwoli to na wyekstrahowanie czystego wpływu interferencji szerokopasmowej na badane sygnały akustyczne. Niemniej, kompleksowe zbadanie rzeczywistych przestrzeni odsłuchowych będzie wymagało uwzględnienia także tych pominiętych elementów.

1.3. Struktura pracy

Prezentacja wyników zaplanowanych celów wraz z opisem szerszego kontekstu analizowanego zagadnienia wymaga odpowiedniej struktury pracy, która jest w dalszej części skonstruowana następująco.

W drugim rozdziale pracy przeanalizowano system złożony z zespołu źródeł szerokopasmowych w kategoriach systemu komunikacyjnego. Omówiono zagadnienie superpozycji sygnałów dochodzących do słuchacza z wielu źródeł, czyli problem interferencji szerokopasmowej, prowadzący do istotnej modyfikacji odbieranych sygnałów. Zniekształcenia wprowadzane przez system nagłaśniający zbadano przyjmując model systemowo-liniowy, w którym właściwości transmisyjne systemów komunikacji są określane funkcjami odpowiedzi impulsowych i charakterystykami częstotliwościowymi (funkcjami przenoszenia). Przestudiowano mechanizm formowania się pola akustycznego dla przypadków: liniowego szyku źródeł, który w obszarach pola odpowiednio oddalonych od apertury umożliwia wyznaczenie regularnych funkcji przenoszenia i charakterystyk kierunkowych oraz dla dowolnego przestrzennego rozmieszczenia elementów promieniujących.

W rozdziale trzecim przedstawiono zastosowaną w niniejszej pracy metodologię badania pola akustycznego. Omówiono wykorzystanie metody odpowiedzi impulsowej, pozwalającej na kompleksowy opis charakteru zniekształceń powstających w obszarze odsłuchu. Opisano sposób uzyskania odpowiedzi impulsowych, z wykorzystaniem metody korelacyjnej z zastosowaniem sygnałów MLS. Zaprezentowano również napisaną przez autora aplikację do wyznaczania wskaźników jakości pola akustycznego. Omówiono pomiary przeprowadzone dla rzeczywistych układów akustycznych.

Rozdział czwarty zawiera podstawową wiedzę z zakresu badania jakości oraz zrozumiałości mowy. Omówiono cel i przebieg procesu parametryzacji sygnałów akustycznych, zarówno w dziedzinie czasu jak i częstotliwości. Zaprezentowano i scharakteryzowano podstawowe parametry czasowe i częstotliwościowe opisujące sygnał mowy. Omówiono również podejście perceptualne, uwzględniające w procesie badania jakości przekazu mowy mechanizm słyszenia ludzkiego ucha. Opisano zastosowanie analizy predykcyjnej LPC i cepstralnej do konstruowania wskaźników pozwalających na ocenę jakości pola akustycznego.

W rozdziale piątym przedstawiono przekrojowo najważniejsze stosowane obecnie metody badania jakości i zrozumiałości sygnałów mowy. Opisano metody wyznaczania miar odległości między wektorami obiektywnych parametrów sygnałów głosowych, jako wskaźników jakości pola akustycznego w układach wielоźródłowych, ze szczególnym uwzględnieniem miar opartych na predykcji liniowej. Omówiono również zasadność wyboru metodyki i zastosowanych miar.

W rozdziale szóstym zaprezentowano przykładowe wyniki otrzymane w trakcie badań symulacyjno-pomiarowych jakości pola akustycznego oraz zrozumiałości przekazu głosowego w systemach wieloźródłowych. Wyniki tych badań wskazują na występowanie niepożądanych efektów filtracji, do których dochodzi w każdym punkcie przestrzeni odsłuchu. Zaprezentowane zostały wykresy map zmienności wskaźników jakości pola akustycznego, dla badanych układów teoretycznych i rzeczywistych.

W ostatnim siódmym rozdziale podsumowano wyniki badań potwierdzających, zdaniem autora, tezy postawione we Wprowadzeniu.

2. POLE AKUSTYCZNE ZESPOŁÓW ŹRÓDEŁ SZEROKOPASMOWYCH

Interferencja jako zjawisko fizyczne jest klasycznie odnoszona do pola fal monochromatycznych. W przypadku dyskretnych źródeł sygnału wąskopasmowego rozmieszczonych regularnie, rozkład amplitud, będący efektem superpozycji fal elementarnych, na przemian konstruktywnej i destruktywnej, tworzy charakterystyczny wzór interferencyjny. Zjawisko interferencji w bardziej ogólnym kontekście, odnosi się także do sygnałów szerokopasmowych, w tym również do sygnałów emitowanych w wielogłośnikowych systemach akustycznych.

W przypadku źródeł monochromatycznych, pobudzanych sygnałem wąskopasmowym, praktyczne jest stosowanie pojęcia charakterystyki kierunkowej. Obliczenia i pomiary wiązki kierunkowej anteny akustycznej są dokonywane dla wybranej częstotliwości emitowanej fali, w obszarze odpowiednio oddalonym od apertury, zwanej strefą daleką. W obszarze bliższym źródłu, noszącym nazwę strefy bliskiej, rozkłady pola mają charakter na tyle złożony, że nie można tam określić funkcji rozkładu o podobnie jednoznacznych własnościach, jak charakterystyka kierunkowa.

Ze względu na naturalną przemienność fali, przesunięcie czasowe między poszczególnymi falami o identycznym przebiegu i nieznacznie różniących się amplitudach, docierającymi do punktu obserwacji, prowadzi do efektów o bardzo złożonym charakterze. Przebieg czasowy ulega zniekształceniu, podobnie jak jego widmo częstotliwościowe, w zależności od położenia słuchacza względem układu źródeł. Każdy punkt odsłuchu staje się wówczas osobną realizacją kanału komunikacyjnego.

2.1. Nagłośnienie jako system komunikacyjny

Wieloźródłowe system nagłaśniające są szczególnego rodzaju szerokopasmowymi, wielokanałowymi, liniowymi systemami komunikacyjnymi, w których wielodrożna transmisja prowadzi do specyficznych zniekształceń interferencyjnych sygnału mowy, mogących w wielu obszarach odsłuchu znacząco utrudnić zrozumienie przekazu.

Na system komunikacyjny składają się cztery klasyczne elementy: nadawca, odbiorca, kanał oraz protokół. Nadajnik i odbiornik są technicznymi interfejsami między, odpowiednio, nadawcą a kanałem oraz kanałem a odbiorcą. W badaniu własności transmisyjnych systemu, rozsądnie jest traktować te interfejsy jako integralną część kanału. Protokół to umówiona forma kodowania informacji, wspólna dla nadawcy i odbiorcy.

Przekaz słowny, także ten naturalny, bez stosowania środków technicznych, jest realizacją komunikacji, w której sygnał mowy, zawierający komunikat zakodowany w szczególny sposób przez nadawcę (mówcę), dociera do odbiorcy (słuchacza) poprzez trudny kanał (przestrzeń

audytorium, do której dźwięk emitowany jest bezpośrednio przez mówcę, przez pojedynczy głośnik bądź przez wiele głośników). Odbiorca dokonuje analizy sygnału i dekoduje komunikat w oparciu o protokół, którego funkcję, w tym przypadku, pełni znany obu stronom język przekazu.

Problemy z bezbłędnym odbiorem komunikatu są w równym stopniu rezultatem wpływu czynników obiektywnych, które można uchwycić badając zniekształcające właściwości kanału transmisyjnego oraz wpływ zakłóceń, co trudnych do jednoznacznego scharakteryzowania czynników subiektywnych, wynikających ze złożoności procesów generacji i odbioru sygnałów mowy. Chodzi tu o dykcję mówcy i słuch słuchacza. Nie bez znaczenia dla możliwości bezbłędnego przekazywania komunikatów w niesprzyjających warunkach, jest zdolność mówcy do jasnego formułowania myśli oraz wspólny kontekst kulturowy obu podmiotów komunikacji.

Przekaz słowny jest redundantny, zawiera wiele nadmiarowej informacji umożliwiającej identyfikację komunikatu, także w niesprzyjających okolicznościach. Słuchacz, w istocie, analizuje i koryguje odbierany przekaz na kilku poziomach, identyfikując poszczególne elementy: fonem (głoski), słowo (sylaba), zdanie, komunikat. Pierwotny proces rozumienia mowy dokonuje się na poziomie fonemów. Analizując mowę na poziomie słów, słuchacz dokonuje korekty ex post nieprawidłowo odebranych głosek, gdyż tylko nieliczne kombinacje fonemów tworzą prawidłowe słowa – taka korekcja jest wyłączona w przypadku zastąpienia słów sylabami pozbawionych znaczenia (logatomami). Podobnie na poziomie zdań może zostać skorygowane nieprawidłowe słowo, jako że nie każde zestawienie słów tworzy prawidłowe zdanie. Z kolei na poziomie komunikatu okazuje się, że nie każdy układ zdań tworzy sensowny przekaz, co umożliwia dalszą korekcję. Tak więc na poziomie słów, zdań i komunikatu efekty pogorszenia zrozumiałości są mniej wyraźne. Słuchacz rozumiejący kontekst może prawidłowo zrekonstruować przekaz nawet wówczas, gdy występują bardzo poważne zniekształcenia, bądź wysoki jest poziom zakłóceń.

2.2. Modele układów akustycznych

W celu badania zjawisk zachodzących w polu akustycznym przyjęto dwa uproszczone modele rzeczywistych wieloelementowych układów nagłośnieniowych. Pierwszy z nich to regularny liniowy szyk źródeł, odpowiadający układowi kolumny głośnikowej[1], a drugi to nieregularny układ źródeł rozłożonych na płaszczyźnie, odpowiadający układowi głośników umieszczonych w suficie. Do analizy pola akustycznego takich układów zastosować można różne modele opisu zachodzących tam zjawisk.

W praktyce projektowej, dla oceny poziomu dźwięku w poszczególnych punktach nagłaśnianego pomieszczenia, stosowany jest zapożyczony z optyki model zakładający, że natężenie (powierzchniowa gęstość strumienia mocy) indywidualnej fali promieniowanej przez

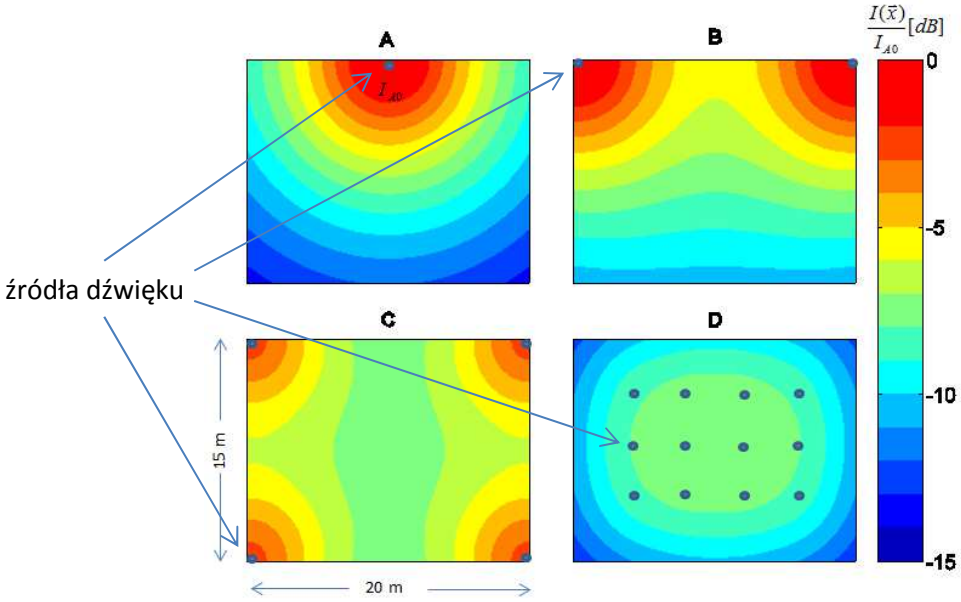
[1] Analizowany tu układ kolumny głośnikowej złożony z kliku identycznych źródeł nie jest tożsamy z zestawem głośnikowym, w którym poszczególne głośniki wysoko- średnio- i niskotonowe są separowane częstotliwościowo poprzez zwrotnicę.

źródło punktowe maleje z kwadratem odległości od źródła, a lokalne natężenie dźwięku można obliczać jako wynik superpozycji natężeń pochodzących z indywidualnych źródeł (głośników). Model ten pomija jednak zjawisko interferencji.

Lepszym podejściem do analizy wielośródłowego pola akustycznego jest badanie sygnałów w oparciu metodę przestrzenno-czasowych odpowiedzi impulsowych. W tym ujęciu w każdym punkcie nagłaśnianej przestrzeni mogą zostać wyznaczone odpowiedzi impulsowe, pozwalające na kompleksową oceną parametrów systemu.

2.2.1. Model optyczny rozkładu natężenia dźwięku

Dla zweryfikowania problemów związanych z formowaniem pola przez systemy nagłaśniające, przedstawiono obliczenia wykonane dla hipotetycznego audytorium o wymiarach 15 m x 20 m, z doskonale pochłaniającymi powierzchniami ograniczającymi, bez wyposażenia (rys. 2.1). Założenie braku odbić na granicach oznacza, że badany model jest, z punktu widzenia zjawisk falowych, fragmentem trójwymiarowej przestrzeni swobodnej. Pozwala to m.in. abstrahować od problemów pogłosu. Audytorium jest nagłośnione na cztery sposoby - jednym, dwoma, czterema i dwunastoma (3 x 4) głośnikami umieszczonymi na wysokości 4 m powyżej płaszczyzny odsłuchu. Moc poszczególnych źródeł dobrano tak, by łączna moc wypromieniowana do obszaru odsłuchu była dla wszystkich konfiguracji jednakowa. Zgodnie z oczekiwaniami, względny rozkład natężeń $I(\vec{x})/I_{A0}$ okazał się najbardziej równomierny dla największej liczby źródeł rozmieszczonych nad płaszczyzną odsłuchu.

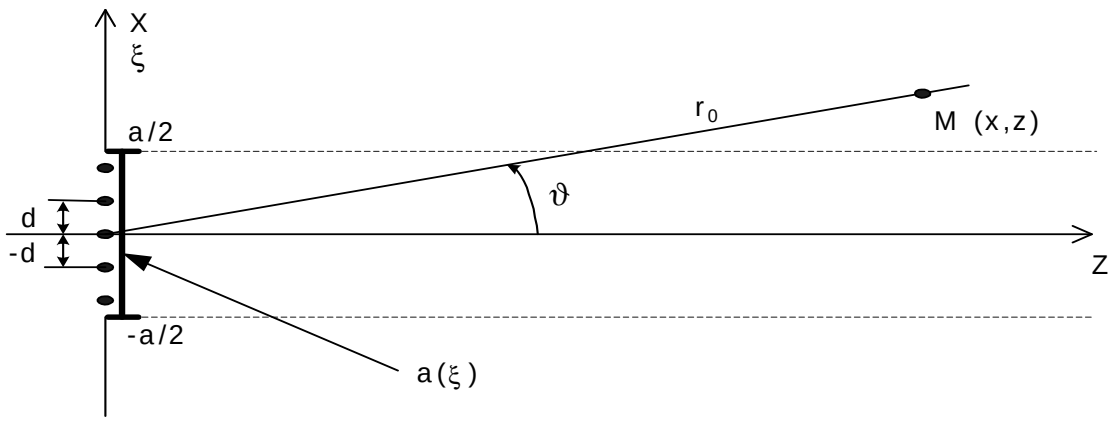


Rysunek 2.1. Rozkład natężenia dźwięku $I(\vec{x})/I_{A0}$ w modelowym audytorium nagłośnionym nieskorelowanymi falami generowanymi przez: A) pojedyncze źródło, B) dwa źródła, C) cztery źródła, D) 12 źródeł.

Model optyczny jest pomocny przy szacowaniu poziomu dźwięku w nagłaśnianej przestrzeni i pozwala prognozować stosunek poziomu pożądanego sygnału do poziomu zakłóceń. Leżące u jego podstaw założenie o dodawaniu natężeń jest słuszne, przez analogię do optyki światła niespójnego, jedynie w sytuacji, gdy szerokopasmowe sygnały promieniowane przez poszczególne źródła są nieskorelowane, co oznacza, że każdy głośnik promieniuje indywidualny sygnał quasi-szumowy, nie mający związku z pozostałymi. Model ten nie umożliwia prawidłowej oceny jakości przekazu, a rozwiązania przyjmowane na jego podstawie mogą okazać się wręcz niekorzystne dla właściwości systemu nagłośnieniowego.

2.2.2. Model liniowy układu szerokopasmowego

Rysunek 2.2 przedstawia jednowymiarową aperturę promieniującą w dwóch wersjach – ciągłej i dyskretnej. W pierwszym przypadku jest to linia promieniująca o długości, a w drugim zespół N = 5 źródeł punktowych rozstawionych co d = a/N. Dla uniknięcia zawiłości formalnych nie mających istotnego znaczenia dla prezentowanych dalej rozważań przyjęto, że punkt obserwacji $M(\vec{x})$ znajduje się w obszarze leżącym na zewnątrz pasa przyosiowego ($|x| > a/2$) [13].



Rysunek 2.2 Geometria układów promieniujących: apertura ciągła a = 1 m i apertura dyskretna N = 5, d = 0,2 m; punkt obserwacji $M(\vec{x})$, $\vec{x} = (x,z) = (r_0, \vartheta)$ [14].

Przy pobudzeniu równomiernym opisanym funkcją aperturową $a(\xi) = rect(\xi/a)$, odpowiedź impulsowa $h(\vec{x},t)$ ma w punkcie M postać opadającej funkcji quasi-hiperbolicznej określonej w granicach czasowych $t \in (t_m, t_M)$, związanych z odległością punktu obserwacji do bliższego i dalszego krańca apertury $|\xi| = a/2$:

$$h(\vec{x},t) = \frac{1}{2\pi} \cdot \frac{1}{\sqrt{t^2 - t_z^2}}, \qquad \text{dla} \quad t \in (t_m, t_M) \tag{2.1}$$

gdzie: $\vec{x} = (x,z)$; $t_z = z/c$, $t_m = r_m/c$, $t_M = r_M/c$

oraz: $r_m = \sqrt{(x - a/2)^2 + z^2}$, $r_M = \sqrt{(x + a/2)^2 + z^2}$.

Transformata Fouriera odpowiedzi impulsowej jest funkcją położenia punktu obserwacji oraz częstotliwości. Oblicza się ją wg wzoru (2.2):

$$H(\vec{x},\omega)=F\{h(\vec{x},t)\}=\int h(\vec{x},t)\exp(-j\omega t)dt \tag{2.2}$$

Funkcję $H(\vec{x},\omega)$ należy interpretować jako uogólnioną funkcję przenoszenia układu apertura promieniująca – obserwator, charakteryzującą jakość transmisji sygnałów szerokopasmowych w polu akustycznym. W przypadku ogólnym obliczenie analityczne tej funkcji nie jest możliwe i konieczne jest zastosowanie numerycznych algorytmów dyskretnej transformaty Fouriera DFT (*ang. Discreet Fourier Transform*).

W strefie dalekiej pola odpowiedź impulsowa $h(r_0,\vartheta,t)$ przyjmuje postać funkcji bramkowej o stałej wysokości:

$$h(r_0,\vartheta,t)=\begin{cases}\dfrac{a}{2\pi r_0}\delta(t-t_0) & dla\ \ \vartheta=0,\\ \dfrac{c}{2\pi r_0|\sin\vartheta|}\ rect\left(\dfrac{t}{t_a|\sin\vartheta|}\right)*\delta(t-t_0) & dla\ \ \vartheta\neq 0,\end{cases} \tag{2.3}$$

gdzie: $\vec{x}=(r_0,\vartheta)$, $t_a=a/c$, $t_0=r_0/c$. Jej transformata Fouriera $H_{r_0}(\vartheta,\omega)$ jest funkcją równocześnie kąta i częstotliwości (czynnik $\exp(-j\omega t_0)$ pomija się dla uproszczenia):

$$H_{r_0}(\vartheta,\omega)=\frac{a}{2\pi r_0}\frac{\sin\left(\dfrac{\omega a}{2c}\sin\vartheta\right)}{\dfrac{\omega a}{2c}\sin\vartheta}=\frac{a}{2\pi r_0}Sa\left(\frac{\omega a}{2c}\sin\vartheta\right) \tag{2.4}$$

co można zapisać też jako:

$$H_{r_0}(\vartheta,\omega)=\frac{a}{2\pi r_0}D(\vartheta,\omega) \tag{2.5}$$

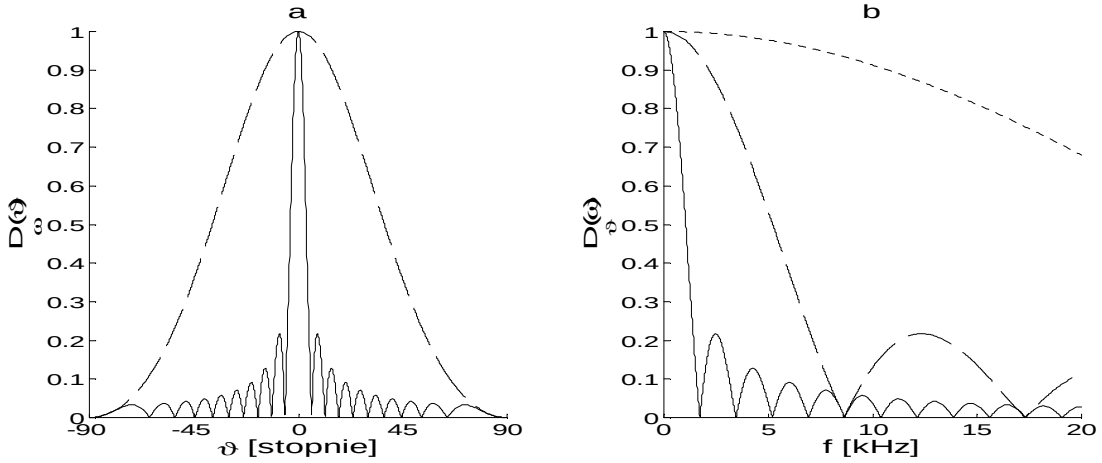
W zależności od potrzeby, bezwymiarową, znormalizowaną funkcję rozkładu kątowo-częstotliwościowego $D(\vartheta,\omega)$ można analizować jako funkcję przenoszenia $D_\vartheta(\omega)$, dla ustalonego kierunku $\vartheta=const$, lub jako charakterystykę kierunkową $D_\omega(\vartheta)$, przy ustalonej częstotliwości $\omega=const$.

Obie funkcje, choć matematycznie niemal identyczne, mają odmienne parametry charakterystyczne, istotne z praktycznego punktu widzenia. Mianowicie, w funkcji $D_\vartheta(\omega)$

występuje parametr ω_0, mający sens częstotliwości odpowiadającej fali o długości λ_0, równej długości rzutu apertury *a* na kierunek obserwacji ϑ (rys.2.3 b):

$$D_{\vartheta}(\omega) = Sa(\pi\omega/\omega_0) \tag{2.6}$$

gdzie: *Sa()* zgodnie z (2.4), $\omega_0 = \frac{2\pi c}{a\sin\vartheta}$, czyli $\lambda_0 = a\sin\vartheta$. Jest to częstotliwość, dla której funkcja przenoszenia w badanym kierunku ϑ ma pierwsze zero. W przypadku apertury o średnicy 1m, pierwsze zero funkcji przenoszenia wypada, odpowiednio, w kierunku $\vartheta = 2°$ - dla częstotliwości f_0 = 43 kHz, w $\vartheta = 10°$ – dla 8,6 kHz oraz w $\vartheta = 60°$ - dla 1,7 kHz. Można to interpretować jako dolnopasmowe własności filtrujące układu promieniującego – im większe jest odchylenie punktu obserwacji od kierunku głównego apertury, tym węższe jest pasmo przenoszenia układu akustycznego.



Rysunek 2.3 a) Charakterystyki kierunkowe $D_{\omega}(\vartheta)$ dla 1,5 kHz (- -) i dla 15 kHz (—) oraz b) funkcje przenoszenia $D_{\vartheta}(\omega)$ dla ϑ= 2° (···), 10° (- -) i 60° (—) w strefie dalekiej pola apertury ciągłej a = 1m [14].

Jak wiadomo, względna średnica apertury $A = a/\lambda$, występująca w funkcji $D_{\omega}(\vartheta)$, określa, dla ustalonej częstotliwości, kierunek ϑ_0 = arcsin $(1/A)$, w którym charakterystyka kierunkowa ma pierwsze zero (rys. 2.3a):

$$D_{\omega}(\vartheta) = Sa(\pi A\sin\vartheta) \tag{2.7}$$

Dla f = 1,5 kHz jest A = 1, co oznacza ϑ_0 = 90°. Dla f = 15 kHz - A = 10 i ϑ_0 = 5,7°.

2.2.3. Systemy wieloźródłowe - zespoły źródeł dyskretnych

Ze względu na geometrię problemu i związaną z nią specyfikę efektów interferencyjnych, systemy wieloźródłowe stosowane w technice nagłośnieniowej można podzielić na dwie kategorie

– układy głośników zwarte i rozproszone. Do pierwszej grupy należą zespoły głośnikowe złożone z kilku identycznych przetworników umieszczonych blisko siebie, z reguły we wspólnej obudowie. Druga grupa, to systemy nagłośnienia sal audytoryjnych, kongresowych i obiektów sakralnych, z licznymi źródłami rozmieszczonymi na ścianach, w suficie bądź na elementach konstrukcyjnych, w których odległość słuchacza od poszczególnych źródeł jest przypadkowa (nawet gdy same źródła są rozmieszczone w sposób regularny).

W przyjętym do analizy problemu modelu systemowo – liniowym sygnał dźwiękowy $s(t)$ niosący przekaz słowny, jest dostarczany do *N* głośników umieszczonych w punktach $\vec{x}_{0i} = (\xi_i, \eta_i, \zeta_i)$. Do słuchacza znajdującego się w punkcie obserwacji $\vec{x} = (x, y, z)$, dociera fala dźwiękowa będąca superpozycją fal pochodzących z indywidualnych głośników (rys. 2.4). Dla uproszczenia pominięto kierunkowość głośników, przyjmując, że mają one własność źródeł punktowych. Dodatkowo założono, że wszystkie kanały akustyczne, włączając głośniki, mają taką samą płaską charakterystykę w całym przenoszonym paśmie częstotliwości.

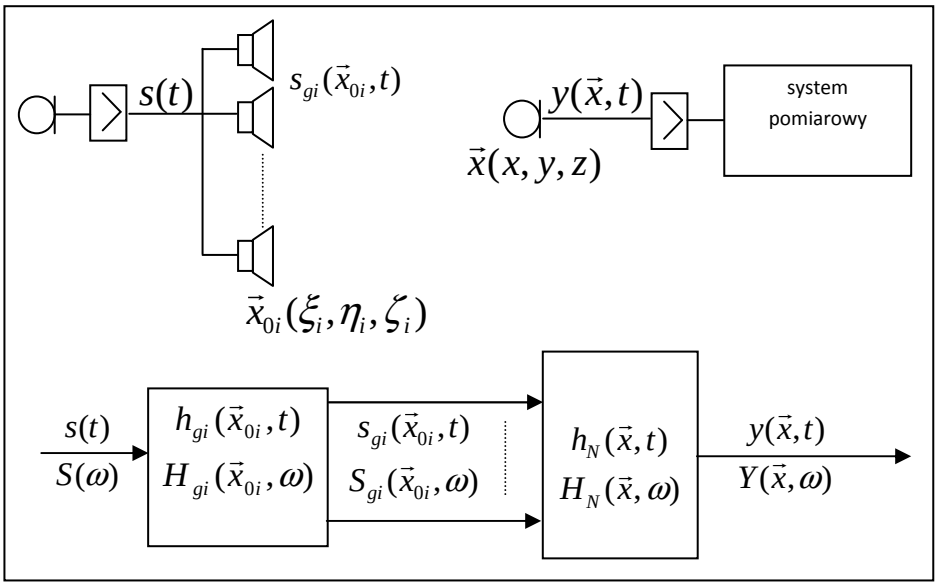
Do analizy sygnałów w polu apertur wykorzystano efektywną metodę przestrzenno-czasowych odpowiedzi impulsowych [15] [6]. Przy tym podejściu w każdym punkcie nagłaśnianej przestrzeni, N-głośnikowy system jest scharakteryzowany przy pomocy funkcji odpowiedzi impulsowej $h_{Ng}(\vec{x}, t)$, która niesie w sobie pełną informację o własnościach transmisyjnych systemu złożonego z kaskady dwóch elementów: głośników jako zespołu źródeł generujących falę dźwiękową oraz przestrzeni przenoszącej falę jako kanału komunikacyjnego, zarówno w stanach przejściowych (transienty), jak i w stanach ustalonych. Funkcja przenoszenia systemu $H_{Ng}(\vec{x}, \omega)$, będąc transformatą Fouriera odpowiedzi impulsowej, charakteryzuje jego własności bezpośrednio w stanach ustalonych.

Własności transmisyjne swobodnej przestrzeni trójwymiarowej opisać można funkcją odpowiedzi impulsowej $h(r, t)$ o symetrii sferycznej. Charakteryzuje ona układ: źródło w punkcie $\vec{x}_{0i} = (\xi_i, \eta_i, \zeta_i)$ – punkt odsłuchu $\vec{x} = (x, y, z)$, znajdującego się w odległości r od źródła, w sposób następujący:

$$h(r, t) = \frac{r_1}{r} \delta(t - r / c) \tag{2.8}$$

gdzie: c [m/s] – prędkość propagacji fali, r_1 = 1 m – odległość jednostkowa od źródła, oraz r [m] - odległość punktu odsłuchu od źródła wynosząca:

$$r = \left[(x - \xi)^2 + (y - \eta)^2 + (z - \zeta)^2 \right]^{1/2} \tag{2.9}$$



Rysunek 2.4 Schemat wyidealizowanego systemu nagłośnienia wielośródłowego (a) i jego model systemowo-liniowy (b) [16].

Jeśli pominąć tłumienie, sama przestrzeń jest częstotliwościowo wszechprzepustowa. O filtrującym wpływie samego kanału akustycznego decyduje dopiero rozmieszczenie źródeł w przestrzeni, w stosunku do miejsca odsłuchu. W systemie N-źródłowym całkowita odpowiedź impulsowa $h_N(\vec{x},t)$ zmierzona w punkcie odsłuchu, ma postać:

$$h_N(\vec{x},t)=\sum_{i=1}^{N} h_i(r_i,t) \tag{2.10}$$

gdzie r_i - odległość i-tego głośnika od punktu odsłuchu równa:

$$r_i=\left[(x-\xi_i)^2+(y-\eta_i)^2+(z-\zeta_i)^2\right]^{1/2} \tag{2.11}$$

W systemie nagłośnieniowym, dodatkowym czynnikiem filtrującym są odpowiedzi czasowo-częstotliwościowe głośników. Właściwości transmisyjne i-tego głośnika opisane są jego funkcją odpowiedzi impulsowej $h_{gi}(t)$. Głośnik pobudzony sygnałem mowy $s(t)$ staje się źródłem fali, której przebieg $s_{gi}(t)$ jest całką splotową w dziedzinie czasu (oznaczoną symbolem $*$) sygnału z odpowiedzią głośnika:

$$s_{gi}(t) = s(t) * h_{gi}(t) \tag{2.12}$$

Odpowiedź impulsowa łańcucha komunikacyjnego złożonego z N różnych głośników ma więc formę:

$$h_{Ng}(\vec{x},t)=\sum_{i=1}^{N}\left[h_{gi}(t)*h_i(r_i,t)\right] \tag{2.13}$$

Jeśli założyć, że głośniki są jednakowe, o odpowiedziach $h_{gi}(t) = h_g(t)$, odpowiedź systemu można zapisać w postaci:

$$h_{Ng}(\vec{x},t) = h_g(t) * \sum_{i=1}^{N} h_i(r_i,t) = h_g(t) * h_N(\vec{x},t) \tag{2.14}$$

Z postaci wzoru (2.8) wynika, że system z jednym źródłem wprowadza jedynie opóźnienie czasowe, jest częstotliwościowo wszechprzepustowy i nie modyfikuje przebiegu sygnału w polu. Natomiast wzór (2.10) opisuje efekty interferencyjne zależne od geometrii źródeł i położenia punktu odsłuchu względem nich. Odpowiadają one za zniekształcenia liniowe sygnału pojawiające się niezależnie od ewentualnych zniekształceń wynikających z nieidealności toru fonicznego i przetworników elektroakustycznych.

2.2.4. Liniowy szyk źródeł dyskretnych

Szczególnym przypadkiem układu wieloźródłowego jest szyk liniowy, w którym źródła są rozłożone równomiernie wzdłuż linii prostej, a odległości między nimi są niewielkie w stosunku do oddalenia punktu odsłuchu (rys. 2.2) [14].

Przy pobudzeniu równomiernym opisanym funkcją aperturową

$$a(\xi) = \sum_{i=-\frac{N-1}{2}}^{\frac{N-1}{2}} \delta(\xi - \xi_i) \tag{2.15}$$

gdzie $\xi_i = id$, odpowiedź impulsowa w punkcie obserwacji $\mathrm{M}(\vec{x})$, ma postać szeregu N impulsów Diraca o hiperbolicznie malejącej wielkości i o rosnących odstępach czasowych, związanych z odległością punktu obserwacji do kolejnych punktów źródłowych. Odstępy te są w ogólności niejednakowe (rys. 2.5 - górny wiersz):

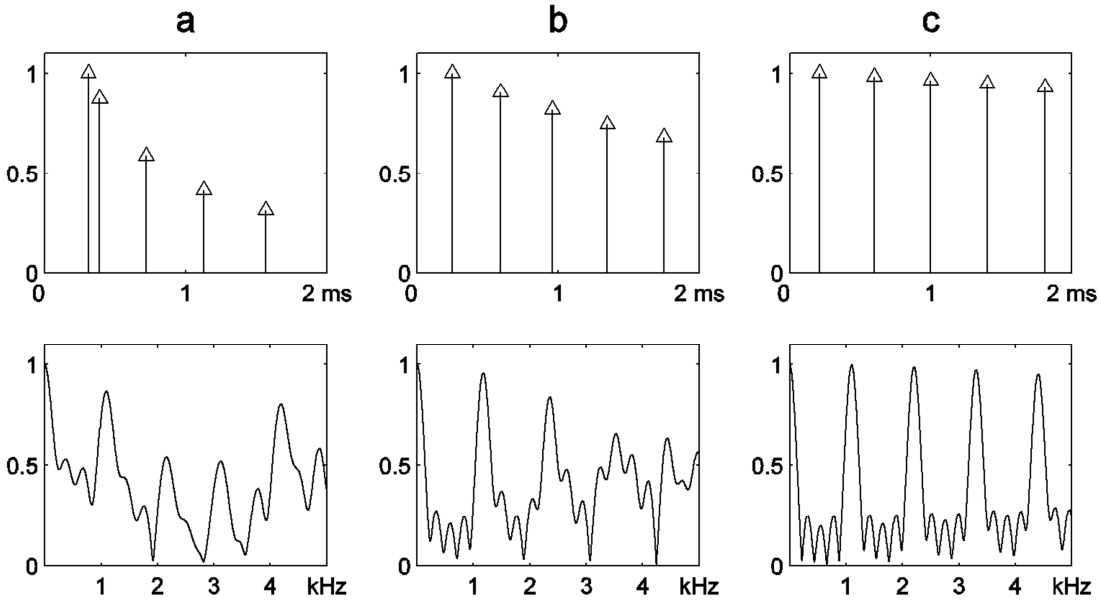
$$h_N(\vec{x},t) = \sum_{i=-\frac{N-1}{2}}^{\frac{N-1}{2}} \frac{r_1}{r_i} \delta(t - t_i) \tag{2.16}$$

gdzie: $t_i = r_i / c$, $r_i = \sqrt{(x - \xi_i)^2 + z^2}$.

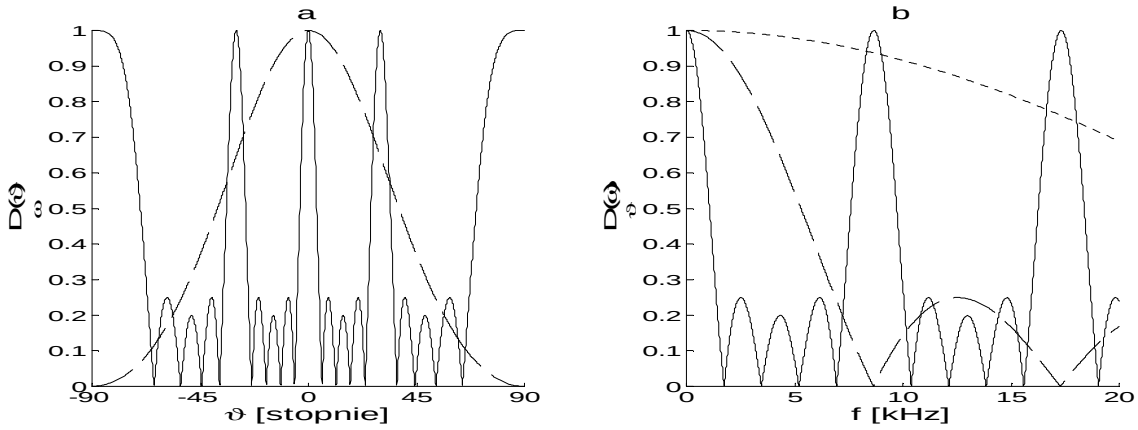
Transformata Fouriera odpowiadająca tego rodzaju funkcji przenoszenia wyraża się w postaci prostej sumy, możliwej do bezpośredniego obliczenia (rys. 2.5 – dolny wiersz):

$$H_N(\vec{x},\omega) = \sum_{i=-\frac{N-1}{2}}^{\frac{N-1}{2}} \frac{r_1}{r_i} \exp(-j\omega t_i) \tag{2.17}$$

Na rysunku 2.5 można zaobserwować ewolucję własności częstotliwościowych układu promieniującego w miarę oddalania punktu obserwacji od apertury. W pobliżu apertury funkcja przenoszenia jest nieregularna (rys. 2.5a), w większej odległości staje się regularna i okresowa (jak w strefie dalekiej - rys. 2.5c).



Rysunek 2.5 Odpowiedzi impulsowe $h_N(r_0, \vartheta, t)$ oraz odpowiadające im funkcje przenoszenia $H_\vartheta(\omega)$ w punktach o współrzędnych ϑ =60°, r_0 = 0.5m (a), 2m (b) i 10m (c). Granica strefy dalekiej r_F wynosi 0.5 m dla f = 0.75 kHz (a), 2 m dla f = 3 kHz (b) i 10 m f = 15 kHz (c) [14].



Rysunek 2.6 a) Charakterystyki kierunkowe $D_\omega(\vartheta)$ dla 1,5 kHz (- -) i dla 15 kHz (—) oraz b) funkcje przenoszenia $D_\vartheta(\omega)$ dla ϑ= 2° (····), 10° (- -) i 60° (—) w strefie dalekiej apertury dyskretnej N = 5, d = 0,2m [14].

W strefie dalekiej, odległości poszczególnych źródeł punktowych do punktu M można uznać, z punktu widzenia czasu propagacji fali, za rosnące jednostajnie ($r_i \approx r_0 - id\sin\vartheta$), przyjmując równocześnie, że związana z tym zmiana wielkości fali jest pomijalna ($1/r_i \approx 1/r_0$). Odpowiedź impulsowa przyjmuje wówczas postać:

$$h_N(r_0,\vartheta,t)=\frac{r_1}{r_0}\sum_{i=-\frac{N-1}{2}}^{\frac{N-1}{2}}\delta\left(t-\frac{r_0-id\sin\vartheta}{c}\right) \tag{2.18}$$

Jeśli obliczyć transformatę Fouriera tej funkcji, pominąć czynnik $\exp(-j\omega t_0)$ i znormalizować, to otrzymuje się kątowo-częstotliwościową funkcję przenoszenia oraz jej przekroje (rys. 2.6) dla ustalonej częstotliwości i dla ustalonego kierunku, w postaci:

$$D(\vartheta,\omega)=\frac{\sin\left(N\omega\frac{d\sin\vartheta}{2c}\right)}{N\sin\left(\omega\frac{d\sin\vartheta}{2c}\right)} \tag{2.19}$$

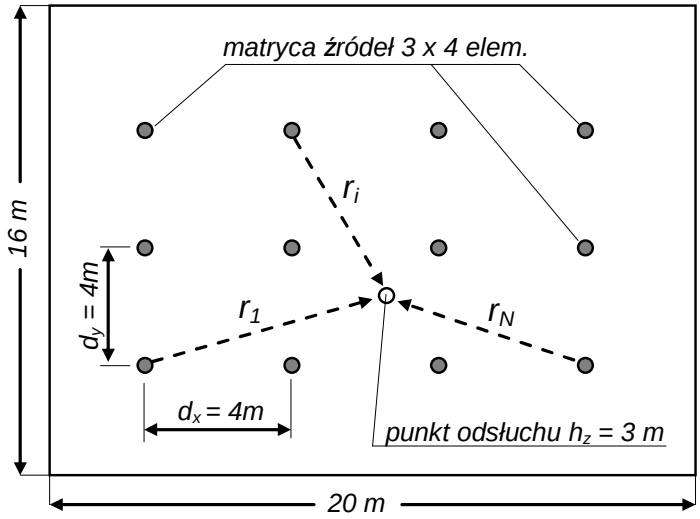
$$D_\omega(\vartheta)=\frac{\sin\left(N\pi D\sin\vartheta\right)}{N\sin\left(\pi D\sin\vartheta\right)} \tag{2.20}$$

$$D_\vartheta(\omega)=\frac{\sin\left(N\frac{\pi}{\omega_0}\omega\right)}{N\sin\left(\frac{\pi}{\omega_0}\omega\right)} \tag{2.21}$$

Rysunki 2.3 i 2.6 ilustrują podobieństwa i różnice pól apertur ciągłych i dyskretnych. Warto zauważyć, że umowna granica $r_F=aA=a^2/\lambda$, od której rozpoczyna się strefa daleka źródła, rośnie wraz z częstotliwością promieniowanej fali. Punkt odbioru może znaleźć się zarówno w strefie bliskiej, jak i dalekiej dla różnych częstotliwości z pasma sygnału (rys. 2.5).

2.2.5. Nieregularny układ źródeł dyskretnych

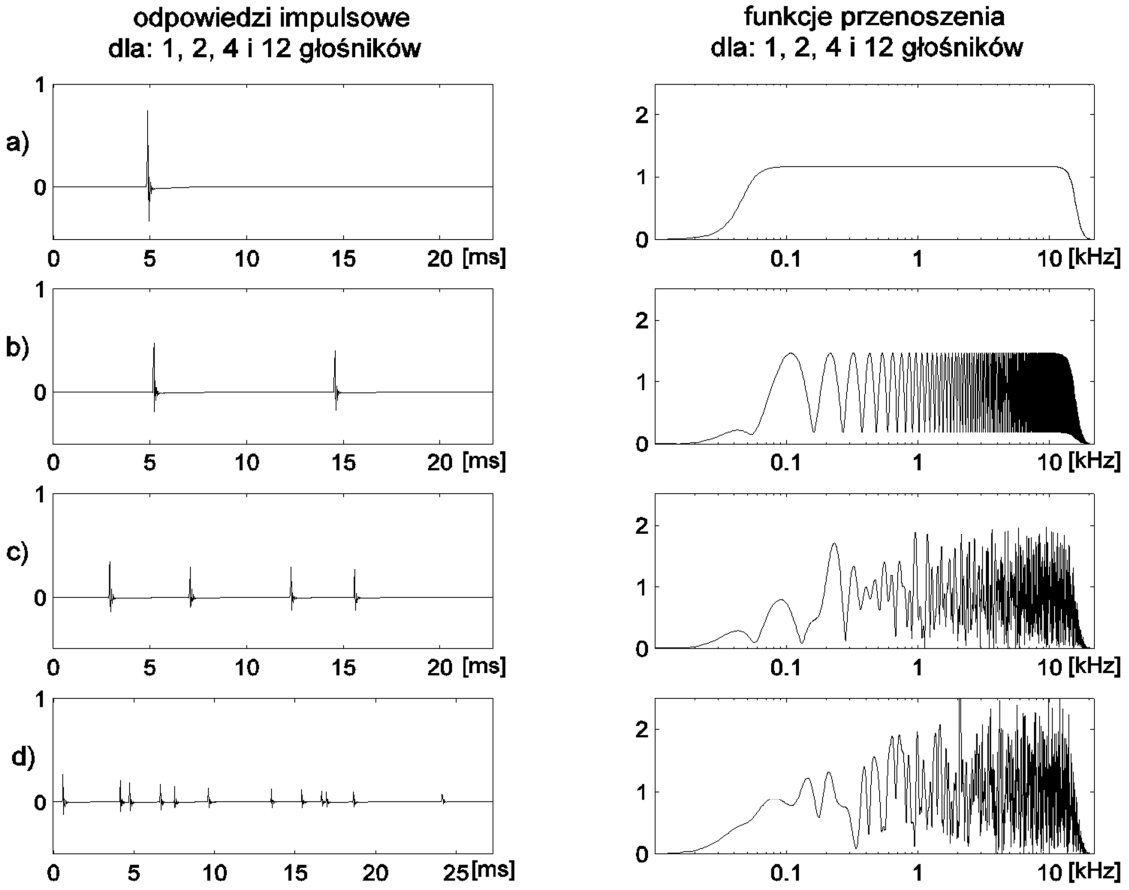
Rozproszone układy źródeł są m.in. wykorzystywane w systemach rozgłoszeniowych, takich jak dźwiękowe systemy ostrzegawcze. Rozmieszczenie źródeł w takich systemach jest często dość przypadkowe i nieregularne względem punktu odsłuchu (rys. 2.7). W ogólnym przypadku, dla wieloelementowych układów promieniujących nie można wyznaczyć analitycznych funkcji opisujących rozkład pola akustycznego i nie jest możliwe operowanie funkcjami typu charakterystyki kierunkowe. Punkt odsłuchu znajduje się zwykle w obszarze pomiędzy elementami promieniującymi i nie istnieje strefa, której można by przypisać np. własności strefy dalekiej pola. W całym zakresie częstotliwości sytuacja odpowiada tzw. strefie bliskiej, w którym to obszarze funkcje przenoszenia są złożone i mocno zróżnicowane w poszczególnych punktach.



Rysunek 2.7 Geometria nieregularnego układu źródeł: głośniki w suficie, odsłuch w płaszczyźnie 3 m poniżej [17].

Rys. 2.8 przedstawia odpowiedzi impulsowe $h_{Ng}(\vec{x},t)$ i odpowiadające im funkcje przenoszenia $H_{Ng}(\vec{x},\omega)$, obliczone dla kolejnych konfiguracji modelu z rys. 2.1 - z jednym, dwoma, czterema i 12 głośnikami, w punkcie odsłuchu. Rysunek 2.9 przedstawia moduł funkcji przenoszenia obliczony dla kolejnych punktów leżących wzdłuż jednej linii płaszczyzny odsłuchu, w stałej odległości od ściany frontowej dla czterech konfiguracji. W każdym przypadku założono, że głośniki są źródłami punktowymi o płaskiej charakterystyce częstotliwościowej w zakresie od 50 Hz do 15 kHz (zobrazowanej na rys. 2.8a wraz z odpowiedzią impulsową modelowego głośnika).

Na rysunku 2.10 przedstawiono przykładowe odpowiedzi impulsowe oraz funkcje przenoszenia wyznaczone w dwóch nieodległych od siebie punktach audytorium z 12-głośnikowym systemem nagłaśniającym. Bardzo nieregularne przebiegi funkcji przenoszenia obliczonych tu jako widma DFT odpowiedzi impulsowych, wskazują na silny efekt filtracji o trudnym do śledzenia charakterze, któremu podlegają transmitowane sygnały, powodujący zniekształcenie dźwięku w każdym punkcie przestrzeni odsłuchu. Dodatkowo, bardzo duże zróżnicowanie kształtu funkcji przenoszenia, mające miejsce nawet w nieodległych punktach, powoduje, że sygnał docierający do lewego i prawego ucha słuchacza jest znacząco inny, co może zaburzyć przestrzenną percepcję źródeł dźwięku.

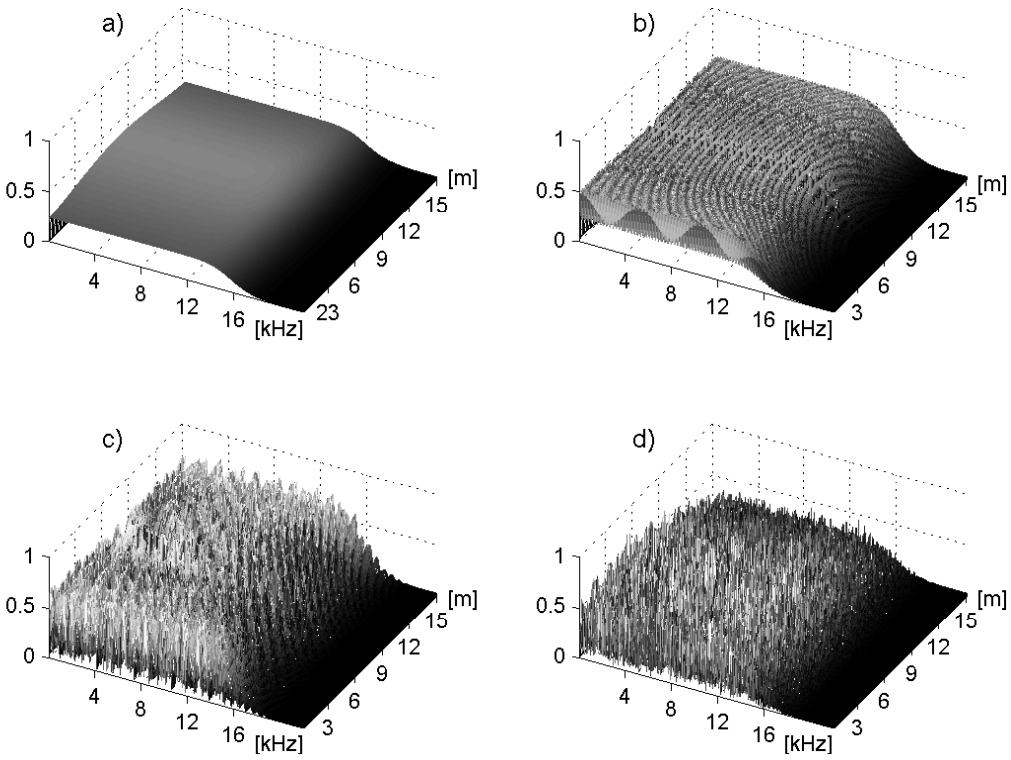


Rysunek 2.8 Odpowiedź impulsowa i funkcja przenoszenia symulowanego systemu nagłośnienia w przykładowym punkcie odsłuchu jak na rys.2.1: a) 1 źródło, b) 2 źródła, c) 4 źródła, d) 12 źródeł; logarytmiczna skala częstotliwości [18].

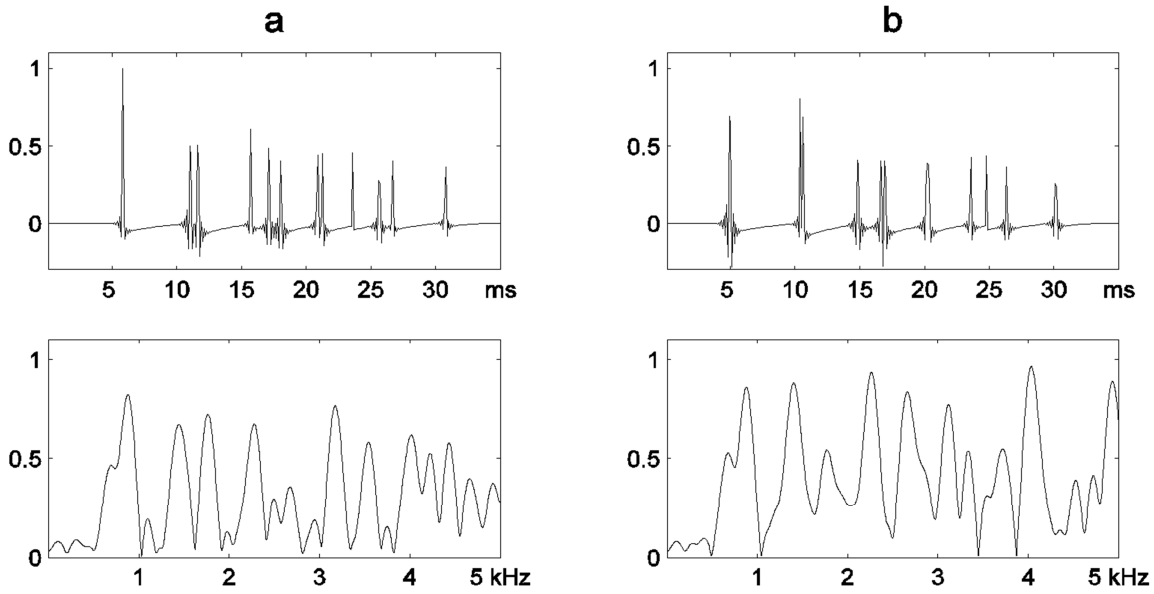
Dla całego pasma sygnału nadawanego, system promieniujący powinien posiadać „płaską" funkcję przenoszenia. Oznaczałoby to brak filtracji i, w konsekwencji, brak zniekształceń liniowych transmitowanego sygnału. Otrzymane wyniki obliczeniowe wykazują jednak na degradujący wpływ zjawiska interferencji szerokopasmowej na właściwości transmisyjne. Warto zauważyć, że kanał akustyczny nie modyfikuje własności systemu z jednym głośnikiem. Natomiast w sytuacji N głośników, odpowiedź impulsowa składa się z N impulsów, które powodują, że funkcja przenoszenia jest bardzo nierównomierna i w każdym punkcie znacząco inna niż w punkcie sąsiednim.

Docierający do słuchacza sygnał $y(\vec{x},t)$ jest splotem sygnału pobudzającego z odpowiedzią systemu:

$$y(\vec{x},t) = s(t) * h_{Ng}(\vec{x},t) \tag{2.22}$$



Rysunek 2.9. Moduł funkcji przenoszenia symulowanego systemu nagłośnienia w punktach odsłuchu wzdłuż linii x = 8.7 m: a) 1 źródło, b) 2 źródła, c) 4 źródła, d) 12 źródeł; liniowa skala częstotliwości [16].



Rysunek 2.10. Odpowiedź impulsowa i funkcja przenoszenia w 2 sąsiednich punktach systemu z rys. 2.7, odległych o 0.2m [17].

2.3. Podsumowanie

Podstawą promieniowania kierunkowego jest interferencja fal w polu akustycznym. Gdy nie ma warunków do jej powstawania nie ma sposobu, by fale promieniować kierunkowo. Idealne interferencje mogą zachodzić tylko dla fal okresowych - różniących się od siebie tyko fazą. Im bardziej złożone (np. szerokopasmowe) sygnały, tym bardziej złożony mechanizm interferencji i mniej wyraźne ich efekty. W skrajnym wypadku sygnałów nieskorelowanych - o charakterze idealnego, białego (o nieskończonym paśmie) szumu, interferencje są doskonale przypadkowe i nie może być efektów kierunkowych.

Jak pokazano powyżej oraz w [14] dla sygnału szerokopasmowego nie da się zdefiniować charakterystyki kierunkowej, gdyż jej kształt zależy silnie od częstotliwości. Można natomiast określić w sposób jednoznaczny, dla wybranego kierunku, pasmo przenoszenia układu akustycznego. Przy zastosowaniu kryterium Rayleigha pasmo to wynosi od zera do $\omega_0/2$, gdzie ω_0 odpowiada częstotliwości, dla której funkcja przenoszenia w badanym kierunku ϑ ma pierwsze zero. Wyznaczając odpowiedzi impulsowych takich układów można określić natomiast ich tzw. uogólnione funkcje przenoszenia, których odpowiednie przekroje mają w strefie dalekiej sens: bądź charakterystyki kierunkowej, dla ustalonej długości fali, bądź częstotliwościowej funkcji przenoszenia, dla ustalonego kierunku promieniowania.

3. BADANIE POLA AKUSTYCZNEGO

Analiza wpływu zjawiska interferencji szerokopasmowej na zaburzenia procesu przenoszenia sygnałów dźwiękowych wymaga przyjęcia odpowiedniej metodyki badania pola akustycznego. Wykorzystanie metody odpowiedzi impulsowej pozwala na kompleksowy opis charakteru zniekształceń powstających w obszarze odsłuchu. Do uzyskania odpowiedzi impulsowych wykorzystano metodę korelacyjną z zastosowaniem sygnałów MLS, zalecaną przez standardy ISO dla akustyki pomieszczeń [19] [20]. Pomiary odpowiedzi impulsowych przeprowadzono dla systemu nagłośnienia w sali audytoryjnej.

3.1. Metoda korelacyjna pomiaru odpowiedzi impulsowej

W korelacyjnej metodzie pomiaru odpowiedzi impulsowych wykorzystuje się specjalne własności szerokopasmowych sygnałów o wąskiej, zbliżonej do impulsu Diraca, funkcji autokorelacji. Gęstość widmowa mocy takiego sygnału testującego jest stała w całym paśmie częstotliwości. Jeśli na wejście systemu podany zostanie odpowiedni sygnał testujący (np. ciąg MLS), funkcja korelacji skrośnej przebiegu wyjściowego oraz wyjściowego będzie dobrą aproksymacją odpowiedzi impulsowej systemu.

W przypadku systemu dyskretnego, odpowiedź y[n] na pobudzenie x[n] ma postać:

$$y[n] = \sum_{m=0}^{L-1} x[m]h[n-m] \tag{3.1}$$

gdzie L jest maksymalną liczbą próbek sygnału x i h (krótszy sygnał uzupełniany jest zerami do długości L):

$$R_{xy}[n] = \frac{1}{L+1}\sum_{m=0}^{L-1} x[m]y[m-n] \tag{3.2}$$

Podstawiając (3.2) do (3.1) otrzymujemy:

$$\begin{aligned} R_{xy}[n] &= \frac{1}{L+1}\sum_{m=0}^{L-1} x[m]\sum_{k=0}^{L-1} h[m]x[m-n-k] \\ &= \frac{1}{L+1}\sum_{k=0}^{L-1} h[k]\sum_{m=0}^{L-1} x[m]x[m-n-k] \\ &= \frac{1}{L+1}\sum_{k=0}^{L-1} h[k]R_{xx}[n-k] \\ &= h[n] * R_{xx}[n] \end{aligned} \tag{3.3}$$

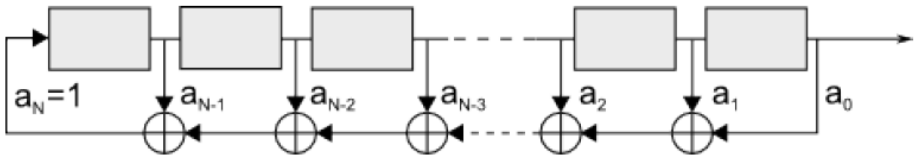
Jeśli funkcja autokorelacji $R_{xx}[n]$ sygnału pomiarowego aproksymuje rozkład delty Kroneckera , funkcja korelacji $R_{xy}[n]$ aproksymuje odpowiedź impulsową systemu:

$$R_{xy}[n] \approx h[n] \tag{3.4}$$

Sygnałami o wąskiej funkcji autokorelacji oraz stałej gęstości widmowej mocy są: biały szum Gaussowski (AWGN – *ang. Additive White Gaussian Noise*), pseudolosowe przebiegi szumowe (PRN – *ang. Pseudorandom Noise*) oraz sygnał szerokopasmowy z modulacją częstotliwości, tzw. chirp. W przypadku sygnału AWGN precyzja pomiaru ograniczona jest szumami ośrodka oraz szumami własnymi systemu pomiarowego. Z tego względu w przypadku metody korelacyjnej bardziej odpowiednie są, łatwe do rekonstrukcji w każdym miejscu procedury pomiarowej, sygnały generowane cyfrowo: cyfrowy sygnał chirp oraz bipolarne przebiegi binarne. Te ostatnie generowane są na bazie pseudolosowych sekwencji MLS, znanych jako sekwencje maksymalnej długości [21] [22].

3.1.1. Ciągi maksymalnej długości

MLS (*ang. Maximum Length Sequence*) jest to okresowy ciąg binarny generowany przez N-stopniowy liniowy rejestr przesuwny ze sprzężeniem zwrotnym – LFSR (*ang. Linear Feedback Shift Register*) - pracujący w konfiguracji Galois'a lub Fibonacci'ego (rys. 3.1). Każdy stan takiego rejestru stanowi liniową kombinację stanu poprzedniego oraz bitu pobieranego z zadanego wcześniej stanu początkowego. Operacje liniowe są realizowane jako funkcje logiczne XOR.



Rysunek 3.1 Rejestr przesuwny w konfiguracji Fibonacciego.

Okresowy ciąg bitów powstaje na wyjściu takiego rejestru, którego postać wielomianowa wygląda następująco:

$$G(X) = X^N + a_{N-1}X^{N-1} + a_{N-2}X^{N-2} + \ldots + a_2X + 1 \tag{3.5}$$

gdzie:

$a_{1\ldots N-1}$ - wagi poszczególnych odczepów rejestrów, przyjmują wartości 0 lub 1,

$X^{0\ldots N}$ - zawartość komórki rejestru, przyjmuje wartości ze zbioru {0,1}.

Długość okresu sekwencji zależy od wartości współczynników a_N, natomiast nie jest zależny od stanu początkowego. Jedyny warunek, który powinien być spełniony to wypełnienie rejestru

tak, by nie były to same zera. Wypełnienie to decyduje jedynie o fazie początkowej ciągu. Maksymalna długość ciągu generowanego przez rejestr o długości *N* wynosi:

$$L = 2^N - 1 \tag{3.6}$$

Sekwencja, której długość wyznacza powyższe równanie (3.6) jest nazywana ciągiem maksymalnej długości (MLS). Dla danej długości N rejestru LFSR istnieje co najmniej jedna kombinacja sprzężeń zwrotnych rejestru, pozwalająca generować ciągi MLS. W przypadku nieprawidłowego doboru odczepów rejestru, generowane będą ciągi o krótszym okresie, niekoniecznie posiadające właściwości pseudolosowe. Takie wielomiany opisujące odczepy z rejestru, aby rejestr generował ciągi MLS nazywane są pierwotnymi. Oznacza to, że nie można doprowadzić ich do prostszej postaci.

3.1.2. Pomiar korelacyjny odpowiedzi impulsowej za pomocą sekwencji MLS

Główna własnością, wykorzystywaną w technikach pomiarowych, jest postać funkcji autokorelacji $R_{xx}[n]$ ciągów MLS, niemal idealnie odwzorowująca pojedynczy impuls:

$$R_{xx}[n] = \delta[n] - \frac{1}{L+1} \tag{3.7}$$

gdzie $\delta[n]$ - delta Kroneckera.

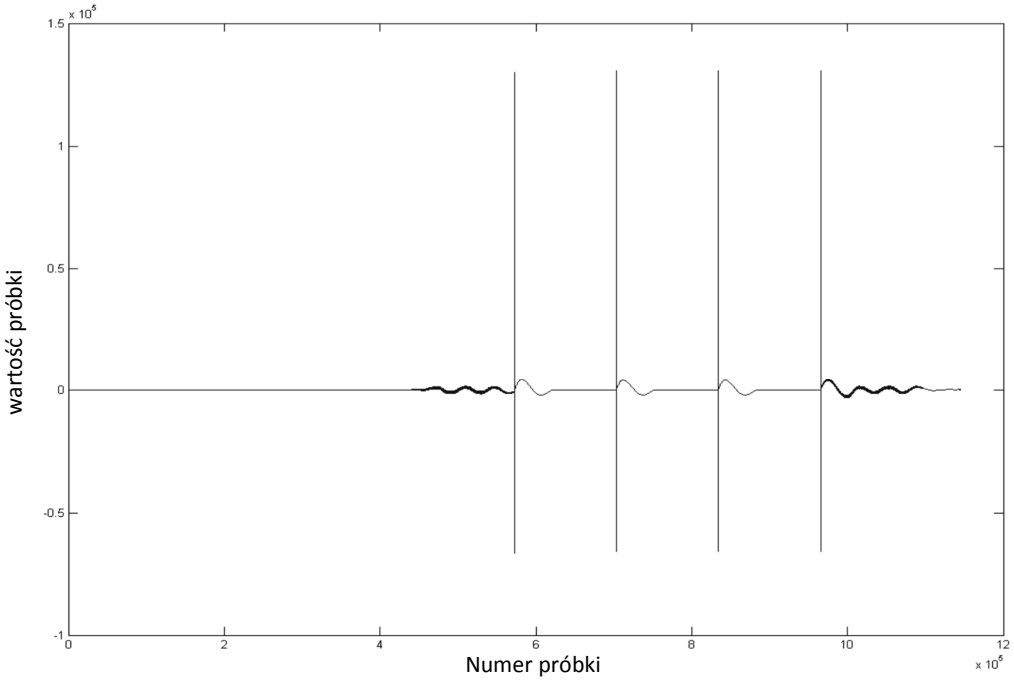
Obecność składowej stałej $\frac{1}{L+1}$ powodowana jest różną liczbą 0 i 1 w generowanym ciągu. Ponieważ ($2^N - 1$) jest liczbą nieparzystą, stąd liczba wystąpień 0 i 1 różni się o jeden. Z tego też powodu im dłuższe ciągi są wykorzystywane, tym mniejsza jest wartość składowej stałej. Podstawiając (3.7) do (3.3), otrzymujemy:

$$R_{xy}[n] = h[n] - \frac{1}{L+1}\sum_{k=0}^{L-1} h[k] = h[n] - \frac{1}{L}\sum_{k=0}^{L-1} h[k] + \frac{1}{L(L+1)}\sum_{k=0}^{L-1} h[k] \tag{3.8}$$

Drugi i trzeci składnik powyższego równania przedstawia średnią składową stałą odpowiedzi impulsowej. Jak wiadomo systemy elektroakustyczne nie przenoszą tej składowej, dlatego może być zaniedbana. Po uproszeniu otrzymujemy zatem:

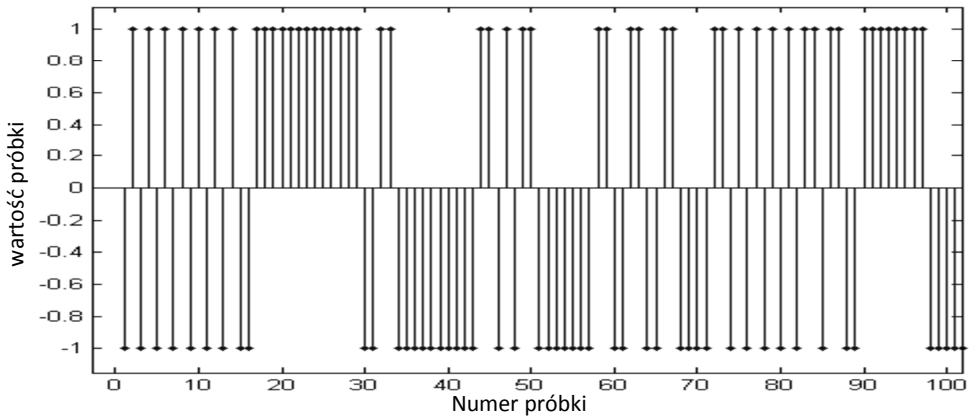
$$R_{xx}[n] = h[n] \tag{3.9}$$

Z równania (3.9) wynika, że przekazując na wejście sygnał MLS, funkcja korelacji skrośnej sygnału wejściowego i wyjściowego daje w przybliżeniu odpowiedz impulsową systemu. Ponieważ sygnał nadawany jak i obliczenia są okresowe, otrzymana odpowiedz impulsowa jest okresowa (rys 3.2), określana jako PIR (*ang. Periodic Impulse Response*) [23].

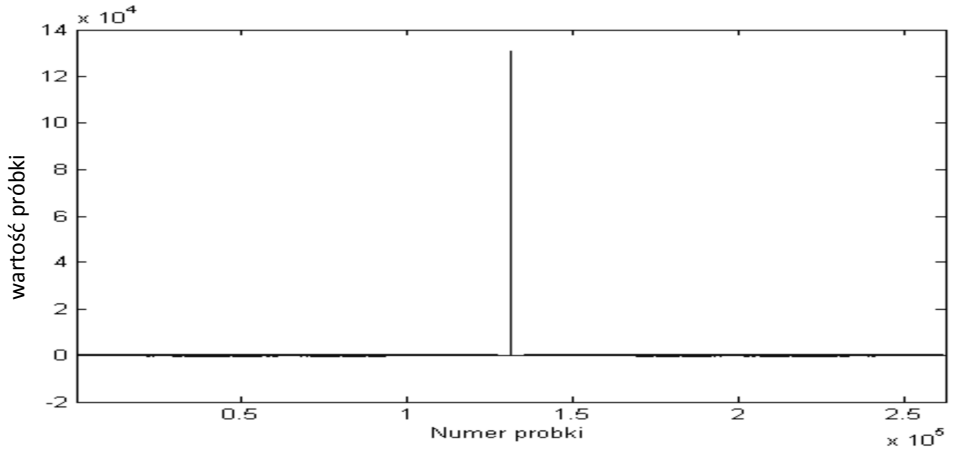


Rysunek 3.2 Przykładowy fragment okresowej odpowiedzi impulsowej (PIR) – MLS rzędu 17 o długości 131071 próbek, powtórzona 4 razy.

Ciągi MLS zazwyczaj generowane są rekursywnie z wykorzystaniem rejestru przesuwnego. Metoda ta, jak już wcześniej wspomniano, wykorzystywana jest również w niniejszej pracy. W wyniku działania takiego rejestru otrzymuje się ciągi wartości binarnych złożonych z 0 i 1. Na potrzeby badań „1” przyporządkowano wartość +1, natomiast „0” wartość -1 tak, by uzyskane wartości rozkładały się symetrycznie względem 0. Poniższy rysunek (rys. 3.3) przestawia przykładowy fragment sygnału MLS.



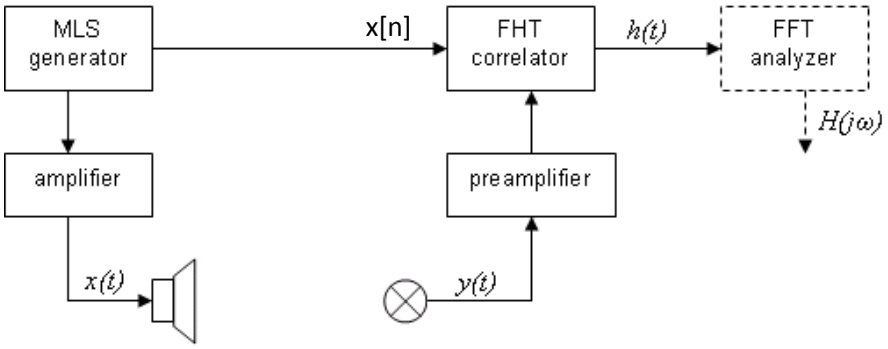
Rysunek 3.3 Przykładowy fragment bipolarnego ciągu MLS .



Rysunek 3.4 Funkcja autokorelacji ciągu MLS rzędu 17.

Wartość $R_{xx}[m]$ dla przebiegów losowych lub pseudolosowych powinna być bliska 0 z wyjątkiem $R_{xx}[0]$, dla którego osiąga wartość $2^N - 1$, czyli wartość maksymalna. Funkcja autokorelacji przybiera wiec postać funkcji delty Kroneckera , co można zaobserwować na rys. 3.4.

Ideę korelacyjnego pomiaru odpowiedzi impulsowych metodą MLS przedstawiono na rys. 3.5. Dyskretny sygnał pomiarowy x[n], którym jest wzmocniona periodyczna sekwencja MLS, po konwersji cyfrowo-analogowej, podawany jest na wejście elektroakustycznego przetwornika nadawczego. Sekwencja MLS powinna być powtórzona co najmniej dwukrotnie. W praktyce stosuje się większą liczbę powtórzeń sekwencji MLS, w wyniku czego otrzymany zostaje ciąg estymat odpowiedzi impulsowej [21].



Rysunek 3.5 Schemat układu do pomiaru odpowiedzi impulsowych metodą korelacyjną MLS [21].

Sygnał y(t) z przetwornika odbiorczego, poddawany jest konwersji analogowo-cyfrowej i wraz z oryginalną sekwencją pomiarową x[n], przetwarzany jest przez algorytm szybkiej transformaty Hadamarda (FHT), który realizuje cykliczny algorytm obliczania funkcji korelacji skrośnej. Wynikowy ciąg liczb reprezentuje dyskretną odpowiedź impulsową systemu liniowego, złożonego z przetwornika nadawczego (np. głośnika), medium propagacyjnego (pomieszczenie) oraz przetwornika odbiorczego (mikrofon akustyczny).

Długość sekwencji oraz częstotliwość próbkowania konwertera cyfrowo-analogowego determinują czas trwania sygnału MLS. Zaletą długiego czasu trwania sekwencji MLS jest większa rozdzielczość funkcji korelacji skrośnej, wadą natomiast jest większy koszt przetwarzania sygnału. Jak wykazano w [24] istnieje podobieństwo permutacyjne między macierzą sekwencji MLS a macierzą Hadamarda. Funkcja korelacji wzajemnej może być obliczana z wykorzystaniem szybkiej transformaty Walsha-Hadamarda, opartej na algorytmie motylkowym, dzięki czemu złożoność obliczeniowa operacji rozplotu sekwencji pomiarowej x[n] oraz odpowiedzi impulsowej h[n] zostaje istotnie zredukowana.

3.2. Pomiary korelacyjne odpowiedzi impulsowych rzeczywistych wieloźródłowych systemów szerokopasmowych

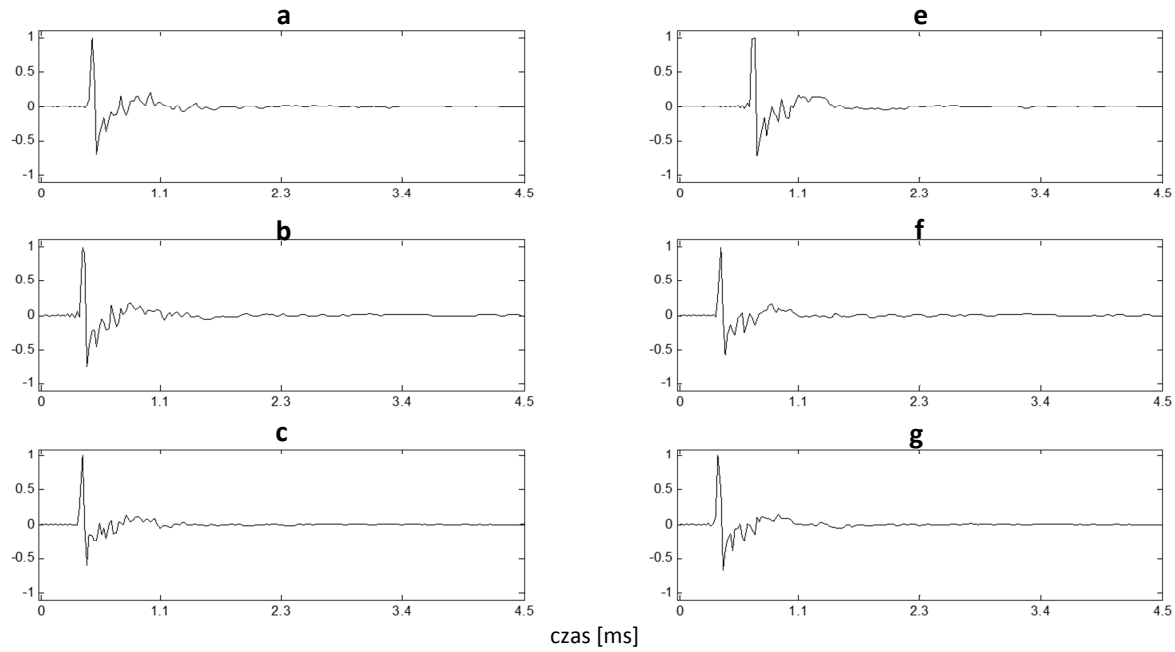
W celu weryfikacji teoretycznych założeń dotyczących istotności wpływu interferencji szerokopasmowej, przeprowadzono pomiary pala akustycznego dla rzeczywistych układów stosowanych do przekazu sygnałów mowy. Aby wyekstrahować efekt przeliczalnej superpozycji sygnałów źródłowych, zastosowano technikę sumowania odpowiedzi impulsowych, wyznaczonych osobno dla każdego ze źródeł. Technika ta pozwala uniezależnić wyniki pomiarów od zakłóceń wywołanych sygnałami pochodzącymi z odbić.

Badano dwa systemy nagłośnienia. Pierwszy z nich to system zainstalowany w sali audytoryjnej, składający się z 12 głośników podwieszonych w suficie, drugi to szyk źródeł złożony z sześciu głośników rozmieszczonych wzdłuż jednej linii. W obu przypadkach wszystkie źródła wchodzące w skład każdego z systemów były niemal identyczne tzn. posiadały bardzo zbliżone częstotliwościowe funkcje przenoszenia. Na rysunkach 3.6 i 3.7 pokazano funkcje przenoszenia oraz odpowiedzi impulsowe sześciu źródeł składowych szyku głośników, pomierzonych w osi głównej w odległości 1 m.

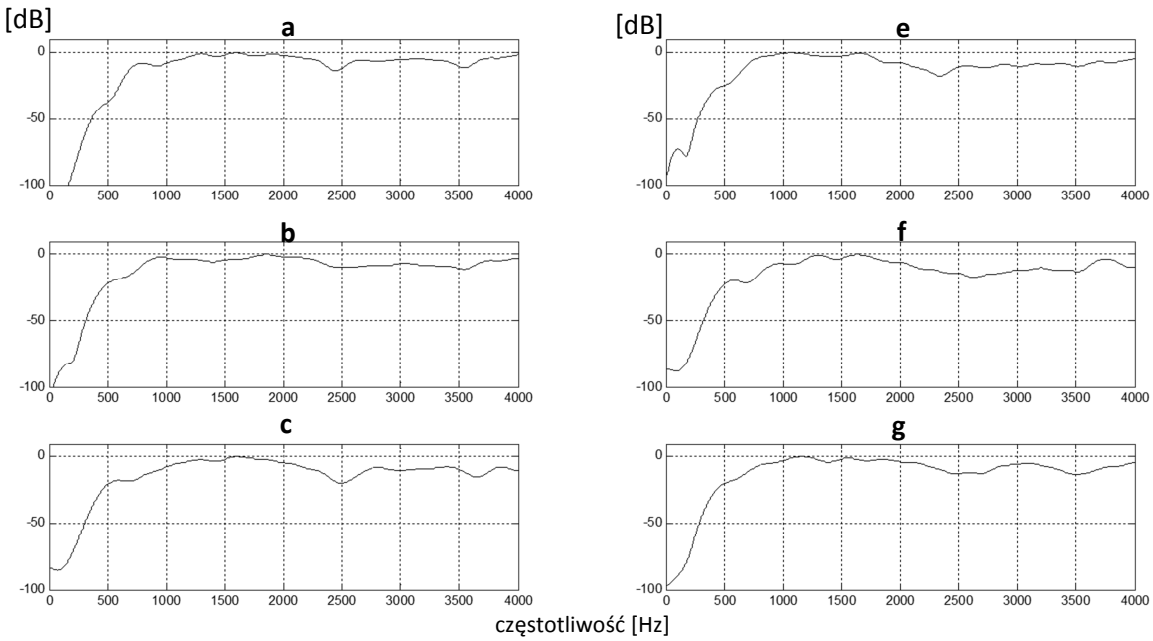
Pomiary odpowiedzi impulsowych systemu nagłośnienia wykonano z użyciem komputera osobistego. Okresowy sygnał MLS rzędu 17, generowany w karcie dźwiękowej z częstotliwością 44,1 kHz, podawano, w miejsce sygnału z mikrofonu mówcy, do wzmacniaczy zasilających zespół głośników. Dźwięk transmitowany przez pomieszczenie był rejestrowany mikrofonem pomiarowym w wybranych punktach odsłuchu, wzmacniany, przetwarzany w przetworniku analogowo cyfrowym na ciąg cyfrowy i podawany do korelatora. Funkcja kroskorelacji ciągu wejściowego i sygnału odbieranego była obliczana z zastosowaniem tzw. szybkiego algorytmu przekształcenia Hadamarda FHT (*ang. Fast Hadamard Transform*) [22].

Ciąg czasowy na wyjściu korelatora odpowiada odpowiedzi impulsowej systemu składającego się z toru elektroakustycznego nadawczego i odbiorczego (wzmacniacz i głośniki plus mikrofon i system pomiarowy z rys. 3.5) oraz N-krotnego kanału akustycznego (przestrzeń między

kolejnymi głośnikami i punktem pomiaru). Przetransformowanie odpowiedzi w dziedzinę częstotliwości przy pomocy przekształcenia Fouriera daje funkcję przenoszenia systemu.



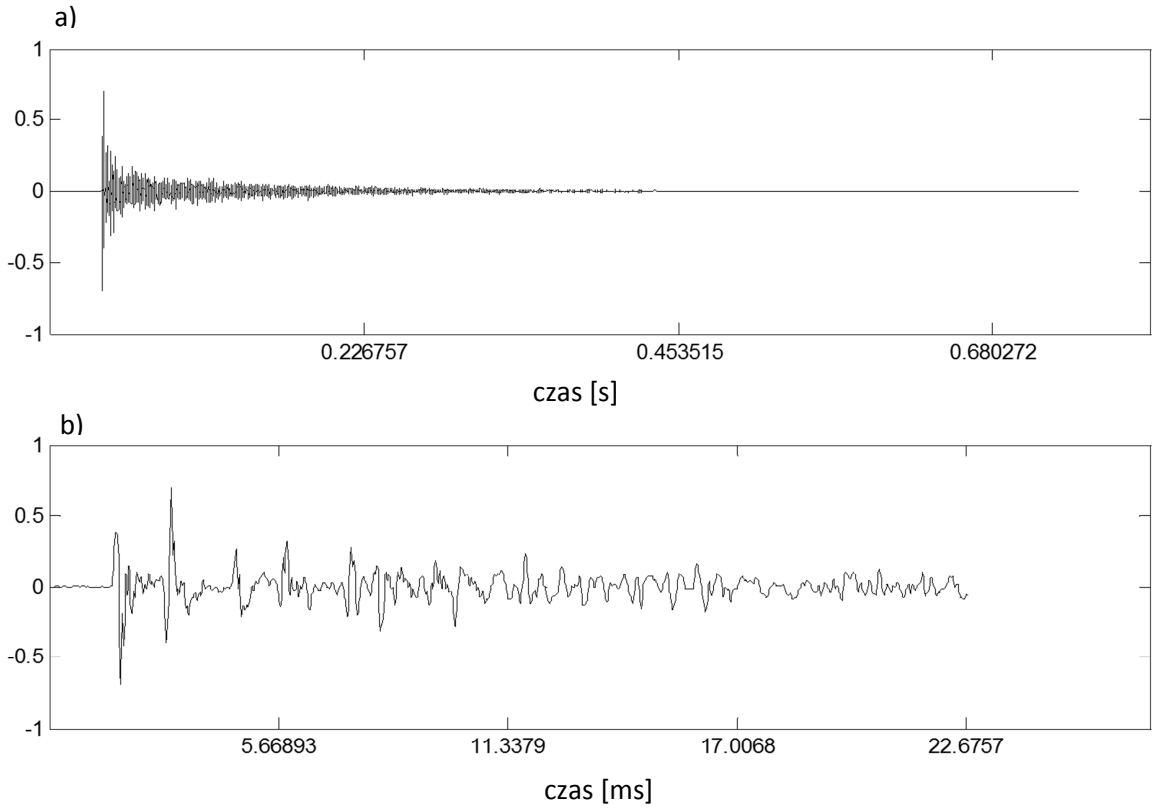
Rysunek 3.6 Odpowiedzi impulsowe sześciu rzeczywistych źródeł (głośników), pomierzone w osi głównej.



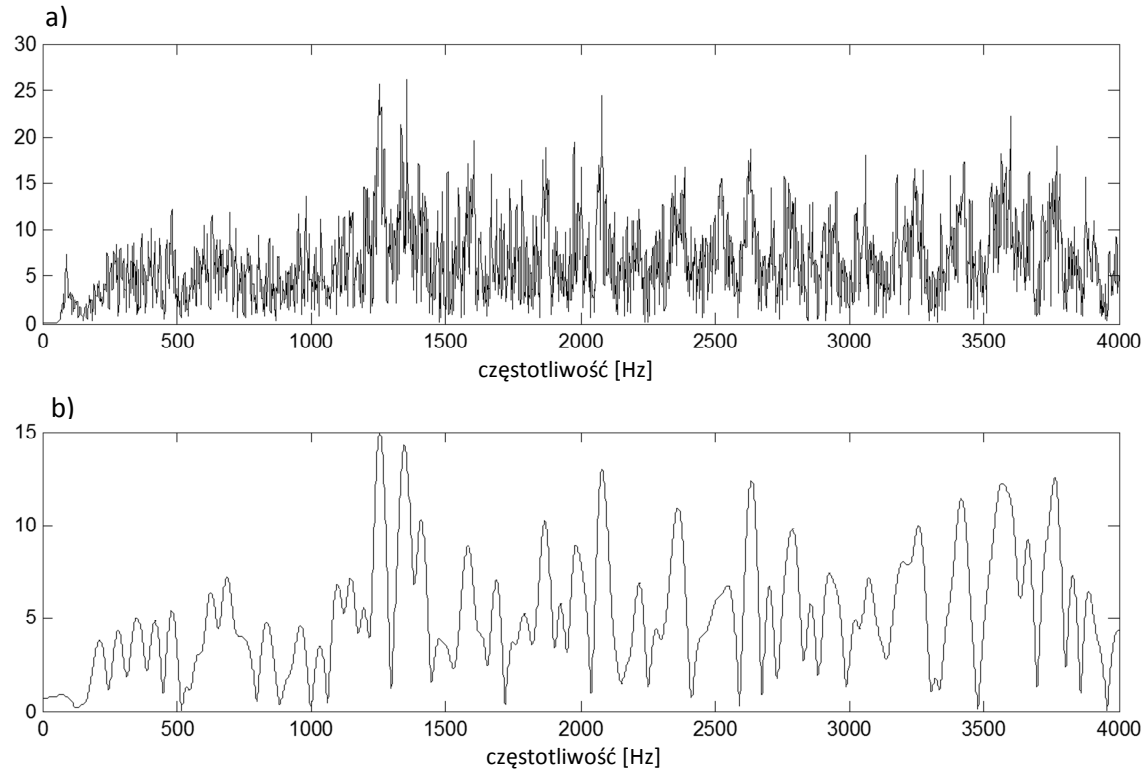
Rysunek 3.7 Częstotliwościowe funkcje przenoszenia sześciu rzeczywistych źródeł (głośników) pomierzone w osi głównej.

Badania systemu nagłośnieniowego sali audytoryjnej przeprowadzono dla 13 punktów pomiarowych, rozmieszczonych równomiernie wśród rzędów siedzeń słuchaczy, w obszarze jednej połówki sali (z uwagi na symetrię pomieszczenia). Na rysunku 3.8 przedstawiono przebiegi czasowe całej (rys. 3.8a) oraz początkowego fragmentu (rys. 3.8b) odpowiedzi impulsowej dla wybranego punktu pomiarowego. Odpowiadające im widma FFT pokazano na rys. 3.9.

Według pomiarów wykonanych dalmierzem laserowym, różnica odległości najbliższego źródła i najdalszego źródła od punktu odsłuchu wyniosła 5.19 m, co przy założeniu prędkości propagacji fali c=340 m/s odpowiada przesunięciu czasowemu t=15.3 ms (rys. 3.8b). Początkowy fragment odpowiedzi impulsowej został więc wybrany tak aby zawierał udziały od wszystkich źródeł systemu.



Rysunek 3.8 Przebieg czasowy odpowiedzi impulsowej w drugim punkcie pomiarowym „P2”. Na górnym wykresie (a) całkowita odpowiedź, na dolnym (b) początkowy fragment z wyraźnymi impulsami pochodzącymi od najbliższych źródeł.

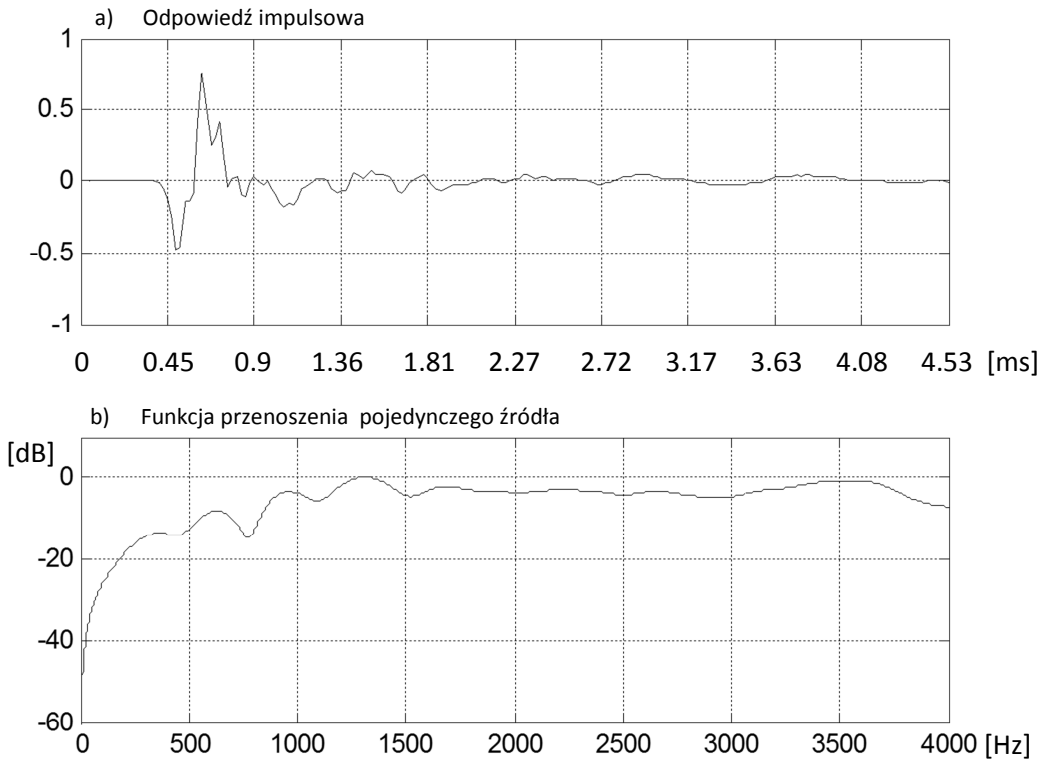


Rysunek 3.9 Widmo częstotliwościowe FFT odpowiedzi impulsowej zmierzonej w drugim punkcie pomiarowym „P2”; a) całej odpowiedzi; b) początkowego fragmentu zawierającego bezpośrednie impulsy od wszystkich źródeł (ok. 20 ms).

Bezpośrednie badania impulsów pochodzących wyłącznie od źródeł jest możliwe jedynie w warunkach pola swobodnego (np. komora bezechowa). Warunki takie można stosunkowo łatwo uzyskać dla pomiarów apertur o niewielkich rozmiarach, np. badany przez autora szyk źródeł rozłożonych w jednej linii (kolumna głośnikowa). Rzeczywisty pomiar systemów nagłośnienia w pomieszczeniach, nawet dla krótkiego przedziału czasu (np. ok. 23 ms jak na rys. 3.8b), jest zawsze superpozycją impulsów bezpośrednich oraz wczesnych odbić od powierzchni ograniczających pomieszczenie, głównie od podłogi. Pierwsze odbicie od podłogi może dochodzi do punktu pomiarowego umieszczonego na wysokości 1.5 m już po ok. 4.5 ms.

W celu zbadania wpływu pochodzącego wyłącznie z bezpośredniego oddziaływania źródeł, autor zastosował syntezę komputerową odpowiedzi impulsowych. Wykonano pomiar odpowiedzi impulsowych osobno dla każdego źródła (rys. 3.10) oraz pomierzono precyzyjnie dalmierzem laserowym odległości wszystkich źródeł od poszczególnych punktów pomiarowych. Na podstawie tych danych zsyntetyzowano całkowite odpowiedzi impulsowe, złożone wyłącznie z udziałów pochodzących od źródeł, uniezależniając się w ten sposób od wpływu wszelkich odbić. Założono jednocześnie, że źródła posiadają dookólne charakterystyki kierunkowe w całym badanym widmie.

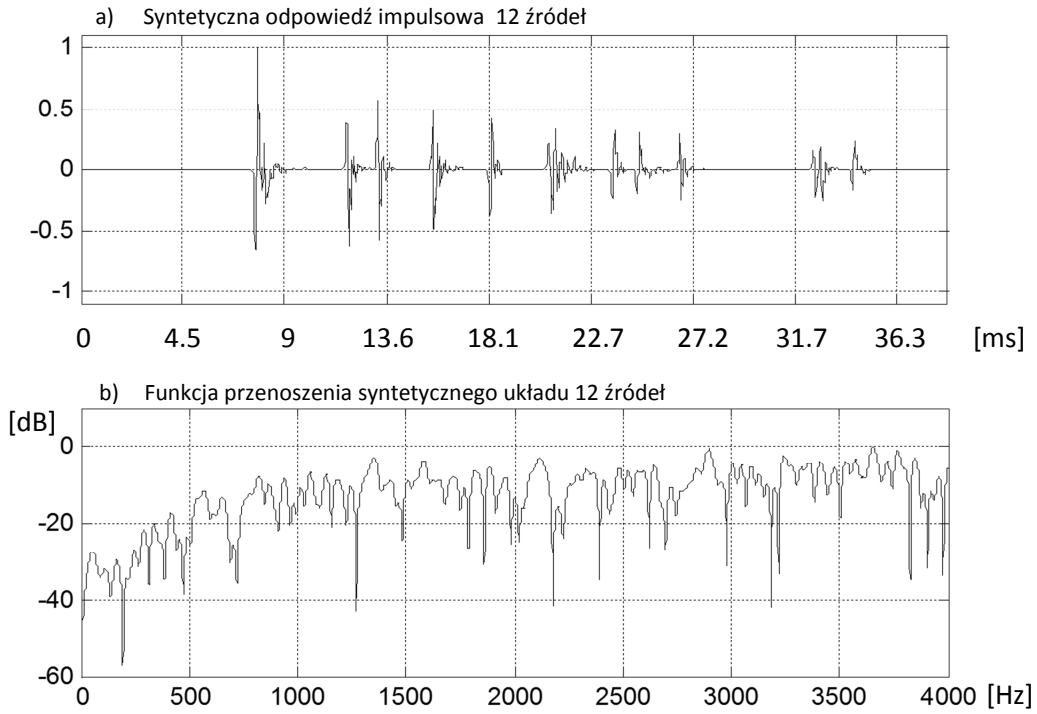
W tabeli 3.1 znajduje się zestawienie odległości punktów pomiarowych od poszczególnych źródeł. Na rysunku 3.11 zaprezentowano przykładową zsyntetyzowaną odpowiedź impulsową dla wybranego punktu pomiarowego.



Rysunek 3.10 Odpowiedź impulsowa oraz funkcja przenoszenia pojedynczego źródła rzeczywistego na podstawie pomiarów w audytorium metodą korelacyjną.

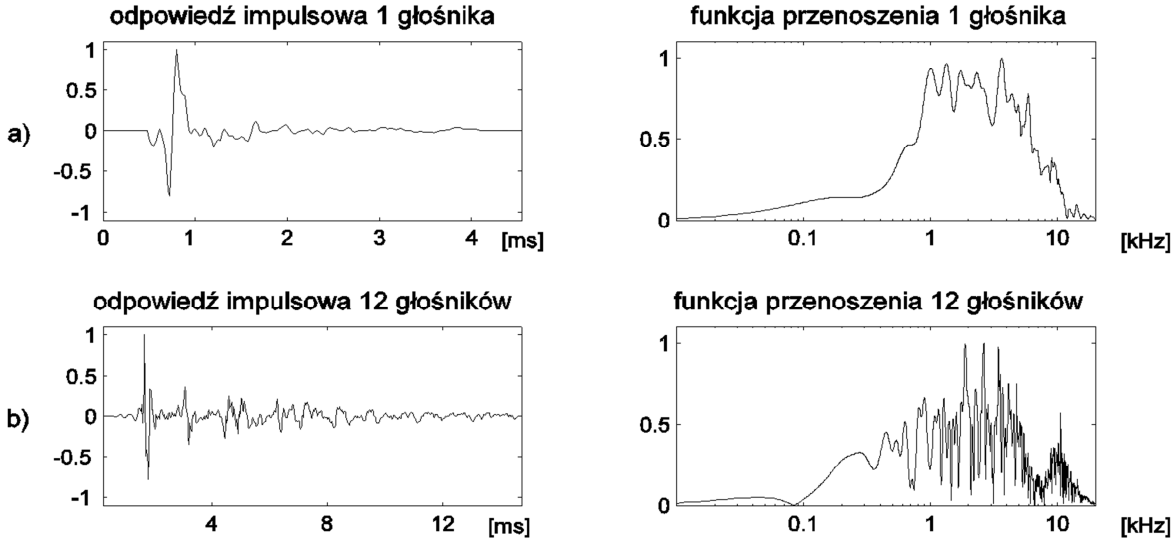
Tabela 3.1 Odległości punktów pomiarowych od poszczególnych źródeł.

Nr źr.	P1 [m]	P2 [m]	P3 [m]	P4 [m]	P5 [m]	P6 [m]	P7 [m]	P8 [m]	P9 [m]	P10[m]	P11[m]	P12[m]	P13[m]
1	2,585	4,136	8,86	9,349	5,638	3,827	4,639	7,775	11,041	10,552	6,21	5,819	3,127
2	4,06	3,746	8,009	8,53	4,11	2,868	4,771	7,632	11,146	9,442	4,447	4,988	3,608
3	6,211	4,814	7,883	8,36	3,725	3,728	5,969	8,271	11,771	8,777	3,572	5,299	5,367
4	4,479	3,169	6,536	6,949	5,391	4,708	2,76	5,054	8,119	8,413	6,301	4,192	2,43
5	5,307	2,743	5,64	6,083	3,808	3,851	3,076	5,183	8,566	7,171	4,594	3,158	3,061
6	7,038	4,178	5,659	6,109	3,415	4,434	4,878	6,322	9,568	6,561	3,727	3,78	5,027
7	7,993	5,138	4,65	4,773	7,044	7,45	3,844	2,636	4,52	6,48	8,061	4,61	5,199
8	8,364	4,794	3,352	3,55	5,835	6,892	4,029	2,945	5,371	4,835	6,725	3,664	5,437
9	9,01	5,613	3,549	3,744	5,482	7,061	5,431	4,755	7,106	4,037	6,027	4,146	6,99
10	11,129	7,932	5,095	4,856	9,452	10,357	6,683	3,793	2,284	6,203	10,381	6,858	8,27
11	11,062	7,397	3,739	3,484	8,245	9,604	6,437	3,795	3,814	4,387	9,063	5,942	7,721
12	11,558	7,655	3,749	3,558	7,675	9,388	7,478	5,202	6,033	3,427	8,254	5,96	8,624



Rysunek 3.11 Zsyntetyzowana odpowiedź impulsową oraz funkcja przenoszenia odpowiadająca parametrom punktu pomiarowego „P1" (tabela 3.1).

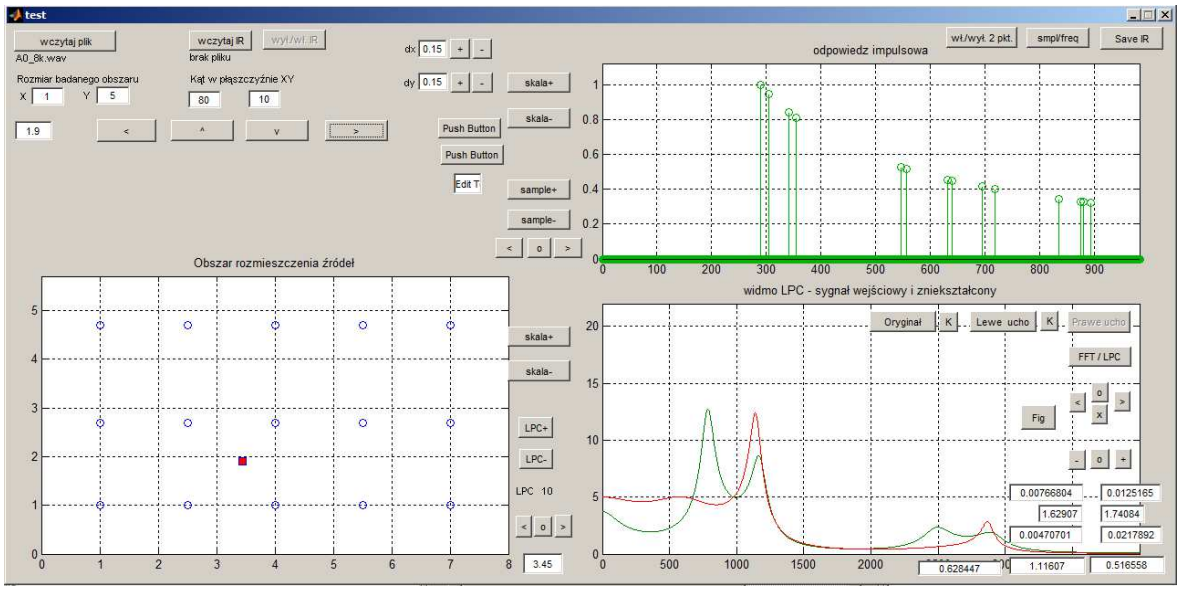
Rysunek 3.12 przedstawia przykładowe wyniki pomiarów przeprowadzonych w audytorium w odniesieniu do pojedynczego głośnika (pomiar wykonany z odległości 1m) oraz całości systemu. W pierwszym przypadku można uznać, że zarejestrowano „czystą" odpowiedź impulsową głośnika, przed dotarciem do punktu pomiarowego sygnałów z pozostałych głośników oraz sygnałów pogłosowych (odbić od powierzchni ograniczających i od wyposażenia - pulpitów, siedzeń). W drugim przypadku mieliśmy do czynienia z pełną odpowiedzią systemu (sygnały z 12 głośników).



Rysunek 3.12 Odpowiedź impulsowa oraz charakterystyka częstotliwościowa (moduł funkcji przenoszenia) w skali logarytmicznej: a) jednego z 12 głośników systemu nagłośnienia sali audytoryjnej oraz b) całego systemu.

3.3. Aplikacja do badania zniekształceń w układach wielośródłowych

Autor zastosował analizę predykcyjną LPC rzędu 10, która umożliwia ekstrakcję do 5 formantów sygnałów głosowych [25], których częstotliwość próbkowania wynosi 8 kHz. Dla oceny zniekształceń, wyznaczano obwiednię widma sygnału testowanego (tzw. pseudowidmo LPC) dla różnych fonemów mowy polskiej, po poddaniu ich splotowi z odpowiedziami impulsowymi systemu nagłośnienia, zarówno obliczonymi dla konfiguracji modelowych, jak i pomierzonymi w rzeczywistym audytorium.



Rysunek 3.13 Widok panelu głównego aplikacji z wczytanym przykładowym układem źródeł oraz testowanym sygnałem mowy (głoska „a”).

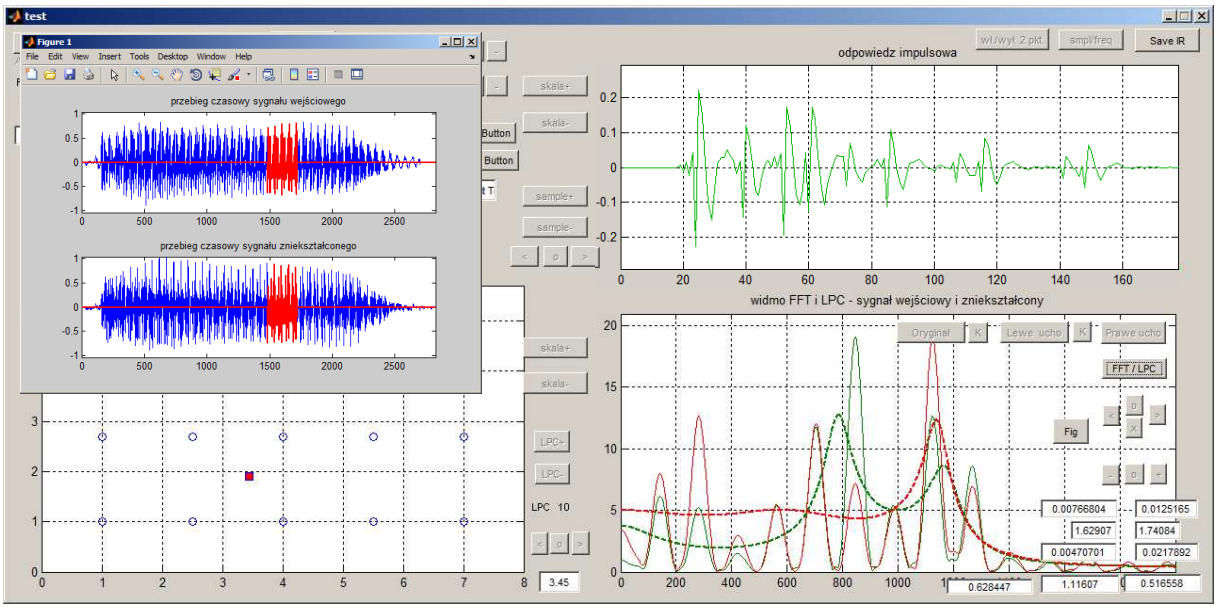
W celu przeprowadzenia wieloparametrowych analiz została utworzona aplikacja (rys. 3.13 i 3.14) symulująca wielośródłowe układy akustyczne która umożliwia:

- Wczytywanie pliku z sygnałem dźwiękowym w formacie WAV (np. głoski).
- Wczytywanie rzeczywistych odpowiedzi impulsowych pochodzących z pomiarów metodą korelacyjną MLS.
- Wprowadzenie dowolnego układu źródeł poprzez wprowadzenie wektorów ich położenia oraz następujących parametrów:
  - rozmiar badanego obszaru (domyślnie 10m x 50m),
  - liczba źródeł wzdłuż osi X i Y (domyślnie 1 x 5),
  - początkowy rozstaw źródeł wzdłuż osi X i Y (domyślnie 1m , 1m),
  - krok rozsuwania/przybliżania źródeł wzdłuż osi X i Y (domyślnie 1m, 1m),
  - środek apertury (x, y) (domyślnie 5m , 38m),
  - początkowe położenia punktu odsłuchu (x, y, z) (domyślnie 5m, 0m, 1m),

- krok przesuwania punktu odsłuchu wzdłuż osi X i Y (domyślnie 0.2m, 1m).

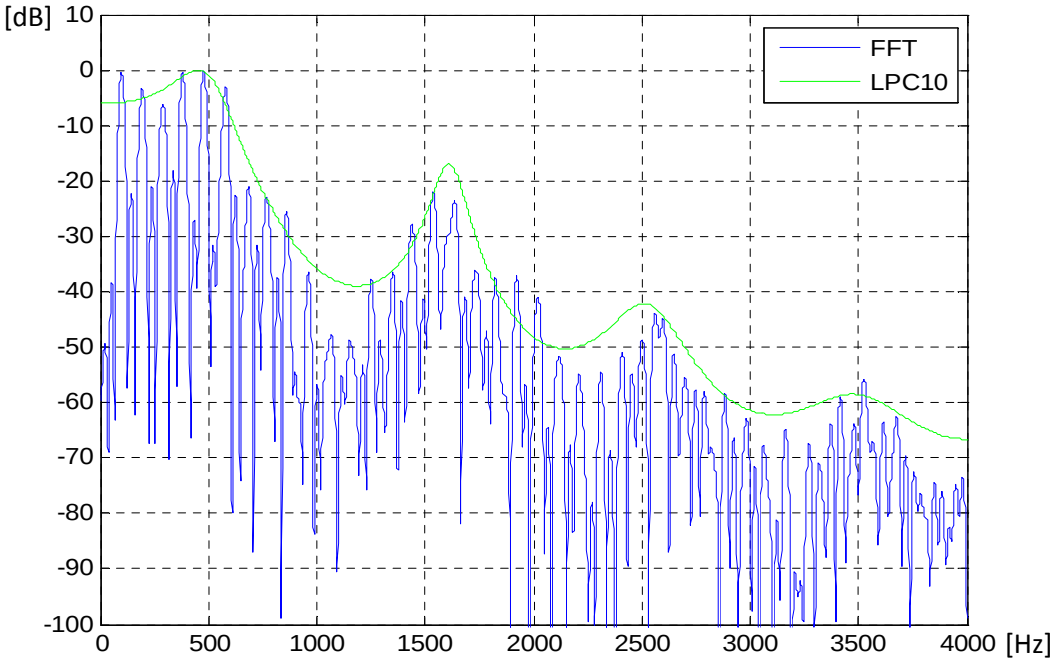
- Wyznaczanie teoretycznych odpowiedzi impulsowych:
  - na podstawie zależności geometrycznych położenia punktu odsłuchu i źródeł wyznaczana jest teoretyczna odpowiedź impulsowa (impulsy o wysokości względnej 1/R),
  - po wczytaniu rzeczywistej odpowiedzi impulsowej wykonywany jest splot z wyliczoną odpowiedzią teoretyczną.
- Dla punktu odsłuchu wyznaczany jest sygnał zniekształcony poprzez splot sygnału oryginalnego, wczytanego z pliku zewnętrznego, z wyznaczoną odpowiedzią impulsową.
  - Zmiana położenia punktu odsłuchu (z zadanym krokiem) wzdłuż osi X oraz Y (wyświetlane są współrzędne punktu odsłuchu oraz kąty odchylenia od osi X oraz Y).
- Zmiana rozstawu źródeł wzdłuż osi X oraz Y (z zadanym krokiem).
  - Dla każdego punktu odsłuchu wyświetlany jest przebieg czasowy odpowiedzi impulsowej z możliwością wyłączania/włączania wczytanej rzeczywistą odpowiedzi impulsowej (po wyłączeniu obliczenia dokonywane są dla odpowiedzi teoretycznej).
- Dla każdego punktu odsłuchu wyznaczane są współczynniki LPC (domyślnie dla rzędu predykcji p=10, możliwość zmiany rzędu predykcji) według następującej procedury.
  - Współczynniki LPC wyliczane są dla fragmentów sygnałów (oryginalnego oraz zniekształconego) o długości l=240 próbek (domyślnie dla sygnałów próbkowanych z częstotliwością 8 kHz).
  - Możliwość zmiany długości fragmentu sygnału, dla którego wyznaczane są współczynniki LPC.
  - Początek analizowanej ramki sygnału domyślnie ustawiony na połowę czasu trwania całego sygnału (możliwość zmiany położenia początkowego ramki).
- W każdym punkcie odsłuchu oraz dla wybranego fragmentu sygnału możliwe jest wyświetlenie przebiegów czasowych sygnału oryginalnego i zniekształconego (z zaznaczonym położeniem analizowanego fragmentu).
- Dla każdego punktu odsłuchu wyświetlane jest pseudowidmo LPC dla sygnału oryginalnego i zniekształconego (widma są skalowane względem oryginalnego w oparciu o stosunek energii sygnału oryginalnego i zniekształconego).
- Dla każdego punktu odsłuchu możliwe jest wyświetlenie widma FFT oraz pseudowidma LPC na wspólnym wykresie (pseudowidma LPC są skalowane względem widma FFT w oparciu o wartość wariancji błędu predykcji oraz długości ramki analizowanego sygnału).

- Dla każdego punktu odsłuchu możliwe jest odsłuchanie:
  - sygnału oryginalnego,
  - sygnału zniekształconego,
  - sygnału oryginalnego po przejściu przez koder LPC; schemat z rys. 3.17,
  - sygnału zniekształconego po przejściu przez dekoder LPC; schemat z rys. 3.18.
- Dla każdego punktu odsłuchu możliwe jest wyeksportowanie do pliku odpowiedzi impulsowej oraz przebiegów widma FFT oraz pseudowidma LPC.
  - Dla każdego punktu odsłuchu wyznaczany jest drugi punkt odsłuchu, który znajduje się w odległości 0.2 m od pierwszego wzdłuż osi X, będący odpowiednikiem drugiego ucha słuchacza. Istnieje możliwość wyświetlenia wszystkich opisanych powyżej charakterystyk dla obu punktów jedocześnie, co pozwala zobrazować rozbieżności w sygnale dochodzącym do obu uszu równocześnie.
- Dla wszystkich wykresów przebiegów odpowiedzi impulsowych oraz pseudowidm LPC aplikacja umożliwia skalowanie na osiach rzędnych i odciętych oraz przesuwanie przebiegów wzdłuż osi odciętych.
- Aplikacja wylicza, zapisuje w pliku oraz prezentuje miary odległości LLR (Log-Likelihood Ratio), IS (Itakura-Saito), CD (Cepstrum Distance), melCD (MFCC Distance) dla każdego punktu pomiarowego.
- Miary odległości prezentowane są w formie liczbowej dla poszczególnych punktów oraz w postaci map zmienności dla całego badanego obszaru (przykładowe mapy zamieszczono w rozdziale 6 oraz Dodatkach)

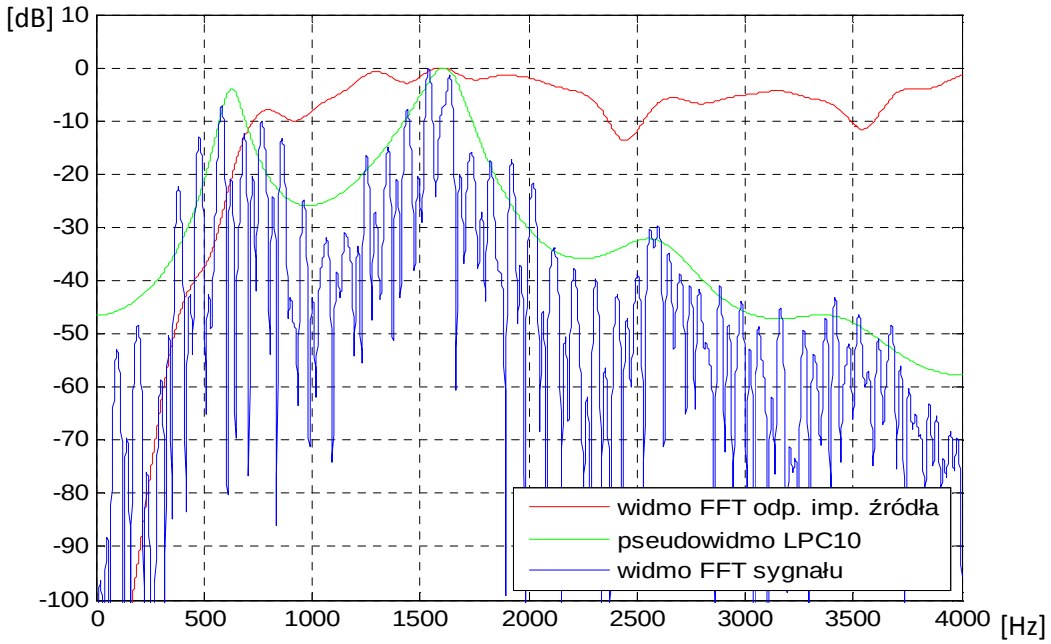


Rysunek 3.14 Widok panelu aplikacji z włączonym podglądem przebiegów sygnałów oryginalnego i zniekształconego oraz wczytaną rzeczywistą odpowiedzią impulsową.

Analizując rzeczywiste układy źródeł należy mieć na uwadze ich wpływ na zniekształcenia charakterystyk częstotliwościowych przenoszonych sygnałów. Zniekształcenia te, są niezależne od rozkładu przestrzennego źródeł, będącego przedmiotem niniejszych badań, a wynikają jedynie z jakości poszczególnych głośników. Na rysunkach 3.15 i 3.16 przedstawiono zmianę charakterystyki częstotliwościowej przykładowego fonemu po wyemitowaniu przez rzeczywiste źródło. Głośnik stanowiący w istocie filtr górnoprzepustowy wpływa istotnie na zmianę poziomu pierwszego formantu.



Rysunek 3.15 Charakterystyki częstotliwościowe dla głoski „e” wyemitowanej przez źródło idealne.



Rysunek 3.16 Charakterystyki częstotliwościowe dla głoski „e” (tej samej co na rys.3.15), wyemitowanej przez źródło rzeczywiste.

3.4. Kodek LPC10

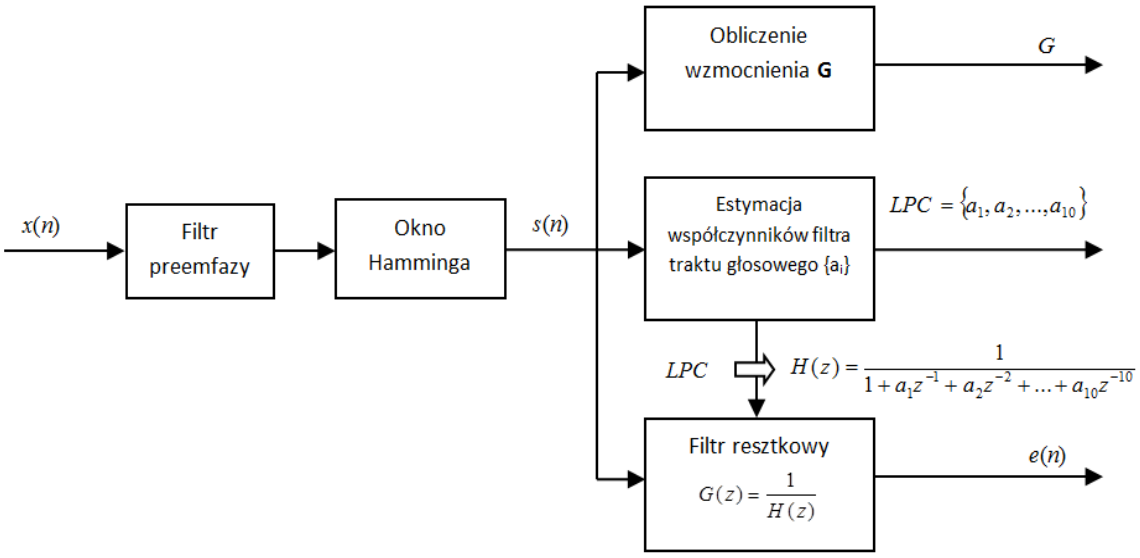
Do aplikacji symulującej wieloźródłowe układy akustyczne wprowadzono również możliwość badania wpływu rozkładu źródeł na jakości przekazu przy zastosowaniu uproszczonego kodeka standardu LPC10. Założono, że:

- zmiany współczynników LPC wpłyną istotnie na poziom zniekształcenia sygnału syntezowanego w dekoderze,
- zniekształcenia wywołane interferencją można przyrównać do błędów w transmisji współczynników LPC, co w efekcie przełoży się na zmianę charakterystyki traktu głosowego.

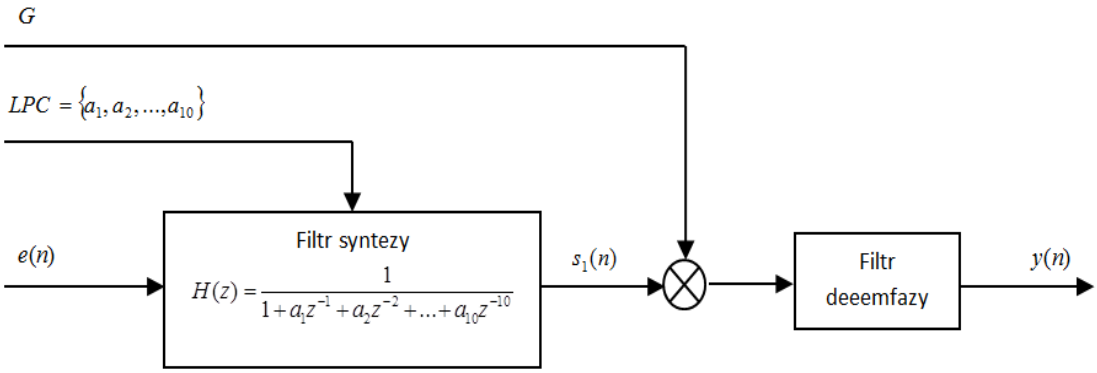
Zaimplementowany w symulatorze kodek z powodzeniem może zostać wykorzystany w subiektywnych testach jakości przekazu mowy, na przykład jako uzupełnienie aplikacji do testów internetowych opisanych w 6.2 [26].

Działanie uproszczonego kodera i dekodera LPC10, którego schemat pokazano na rysunkach 3.17 i 3.18 oparto na następujących założeniach:

- sygnał wejściowy próbkowany jest z częstotliwością f_{pr}= 8000 Hz,
- sygnał analizowany jest w oknie czasowym 30 ms (240 próbek), wymnażanym z oknem Hamminga o długości 240 próbek, przesuwanym co 120 próbek,
- dla każdego okna wyznaczane są parametry, przesyłane do dekodera: współczynniki filtra traktu głosowego, wzmocnienie oraz sygnał pobudzający,
- jako pobudzenie filtra syntezy wykorzystano tzw. sygnał resztkowy, wyznaczony za pomocą filtra odwrotnego do filtra traktu głosowego (tzw. filtr resztkowy); sygnał ten ma mniejszą dynamikę niż sygnał oryginalny,
- na wejściu kodera znajduje się filtr preemfazy, służący do kompensacji tłumienia wyższych częstotliwości, gdyż w rzeczywistym trakcie głosowym, wielokrotności tonu podstawowego są tłumione z nachyleniem ok. 12 dB/oktawę przez rezonatory układu oddechowego,
- zastosowano filtr deemfazy, który jest filtrem odwrotnym do filtra preemfazy.



Rysunek 3.17 Schemat kodera LPC.



Rysunek 3.18 Schemat dekodera LPC.

3.5. Podsumowanie

Badania metodą odpowiedzi impulsowych pola akustycznego uformowanego w układach wieloźródłowych wykazały, iż funkcja przenoszenia takich systemów jest istotnie różna w różnych punktach mierzonego audytorium. W wyniki przeprowadzonych pomiarów uzyskano bazę rzeczywistych odpowiedzi impulsowych odpowiadających punktom odsłuchu rozłożonym w całym obszarze pola. Dane te zostały wykorzystane do badań symulacyjnych różnorodnych układów akustycznych opisanych w rozdziale 6 niniejszej pracy.

4. WPŁYW INTERFERENCJI SZEROKOPASMOWEJ NA PARAMETRY SYGNAŁU MOWY

Podstawową ideą związaną z badaniami jakości mowy jest założenie, iż sygnał mowy można opisać za pomocą zestawu cech umożliwiających ich porównanie. W wyniku parametryzacji otrzymujemy wskaźniki, bądź wektory wskaźników, które w obiektywny sposób opisują cechy sygnałów mowy. W wielu przypadkach możliwe jest bezpośrednie porównanie tych wektorów, dla sygnałów oryginalnych oraz zniekształconych, a wielkość różnicy pomiędzy wartościami parametrów określa stopień zniekształcenia.

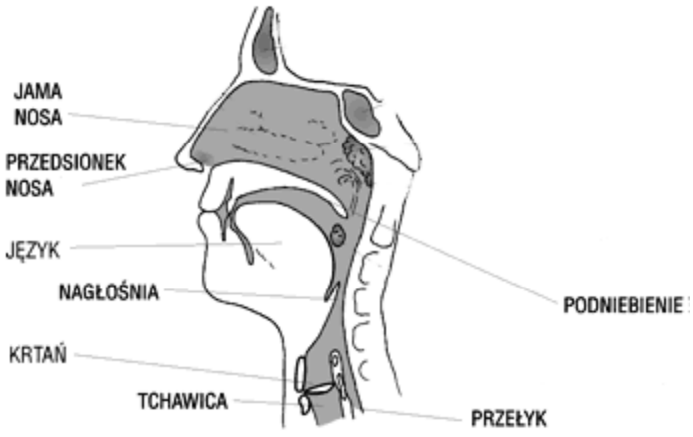
Dla obiektów należących do klasy sygnałów akustycznych wyróżnia się dwa podstawowe kategorie parametrów: czasowe i widmowe (częstotliwościowe). Dla każdej z tych kategorii zdefiniowanych zostało wiele parametrów opisujących cechy sygnałów akustycznych [27]. W kontekście badań nad jakością i zrozumiałością przekazu głosowego szczególnie przydatne okazują się charakterystyczne parametry widmowe – formanty, wyznaczane w oparciu o predykcję liniową.

4.1. Model wytwarzania sygnału mowy – formanty

Proces powstawania sygnału mowy można przedstawić za pomocą modelu składającego się ze źródła (sygnał pobudzający) oraz filtru formującego charakterystykę amplitudowo-częstotliwościową sygnału pobudzającego, w wyniku czego powstają charakterystyczne dla danej głoski lokalne maksima (formanty), decydujące o jej rozpoznawalności.

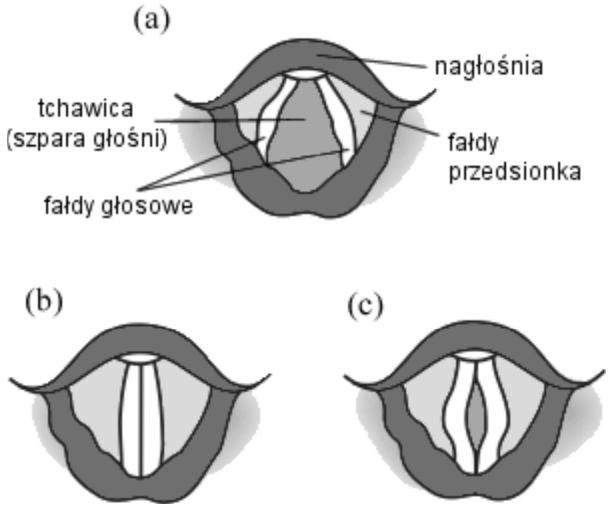
4.1.1. Generowanie sygnału mowy

Mowa powstaje w procesie artykulacji czyli wspólnej pracy wszystkich narządów mowy, w wyniku której wytworzone zostają dźwięki emitowane następnie w postaci sygnału mowy. Do podstawowych narządów mowy zaliczane są m.in.: tchawica, krtań, więzadła głosowe oraz elementy traktu głosowego: jama gardłowa, jama ustna, jama nosowa, język i usta (rys. 4.1).



Rysunek 4.1 Budowa narządu głosowego w przekroju.

Jednym z podstawowych organów uczestniczących w wytwarzaniu dźwięków mowy jest krtań, będąca końcowym fragmentem tchawicy. Krtań, która jest odcinkiem drogi oddechowej, składa się z sześciu chrząstek połączonych z kością gnykową, tchawicą i między sobą więzadłami. Cześć jednego z więzadeł, łącząca chrząstkę tarczowatą z pierścieniową, tworzy tzw. więzadła głosowe, które ograniczają szparę głośni. Mięśnie wewnętrzne krtani rozszerzają i zwężają tą szparę (rys. 4.2) utrudniając przepływ wydychanego powietrza, co wprawia więzadła głosowe w drgania. Czynność głosowa jest więc wynikiem drgania więzadeł głosowych zwanych inaczej fałdami głosowymi. Wysokość głosu zależy od długości fałdów głosowych, ich napięcia, częstości drgań i ciśnienia wydechowego powietrza. Barwa głosu uzależniona jest od budowy gardła, jamy nosowej i częściowo zatok przynosowych.



Rysunek 4.2 Przekrój poprzeczny głośni: (a) stan milczenia, (b) stan mówienia, (c) stan szeptania.

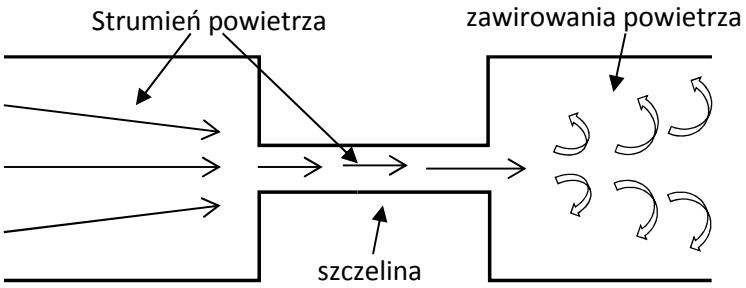
Człowiek może sam świadomie regulować ruchy krtani w celu wygenerowania sygnału mowy. Ruchy te dodatkowo wiążą się bezpośrednio z ruchami żuchwy, języka i podniebienia miękkiego. Właściwa praca mięśni krtani oraz języka, żuchwy itd. są niezbędnymi czynnikami warunkującymi zrozumiałość mowy. Każde ciało drgające posiada własną częstotliwość rezonansową drgań, tzw. ton własny. Ton krtaniowy wzmacniany jest rezonansowo w wnękach powietrznych krtani, kanału nosowego i jamy ustnej. Głos bez tego rezonansu jest matowy i głuchy.

Ton podstawowy (krtaniowy) jest bardzo ważnym parametrem charakteryzującym przebieg sygnału akustycznego mowy. Częstotliwość tonu krtaniowego można wyznaczyć poprzez zastosowanie odpowiedniej filtracji dolnoprzepustowej. Ton podstawowy nazywany formantem zerowym F_0 występuje tylko dla głosek dźwięcznych i jest charakterystyczny dla każdego mówcy oraz rodzaju głosu (tabela 4.1).

Tablica 4.1 Zakresy częstotliwości dla tonu podstawowego (krtaniowego).

Rodzaj głosu	**Częstotliwość F_0**
bas	80-320 Hz
baryton	100-400 Hz
tenor	120-480 Hz
alt	160-640 Hz
mezzosopran	200-800 Hz
sopran	240-960 Hz

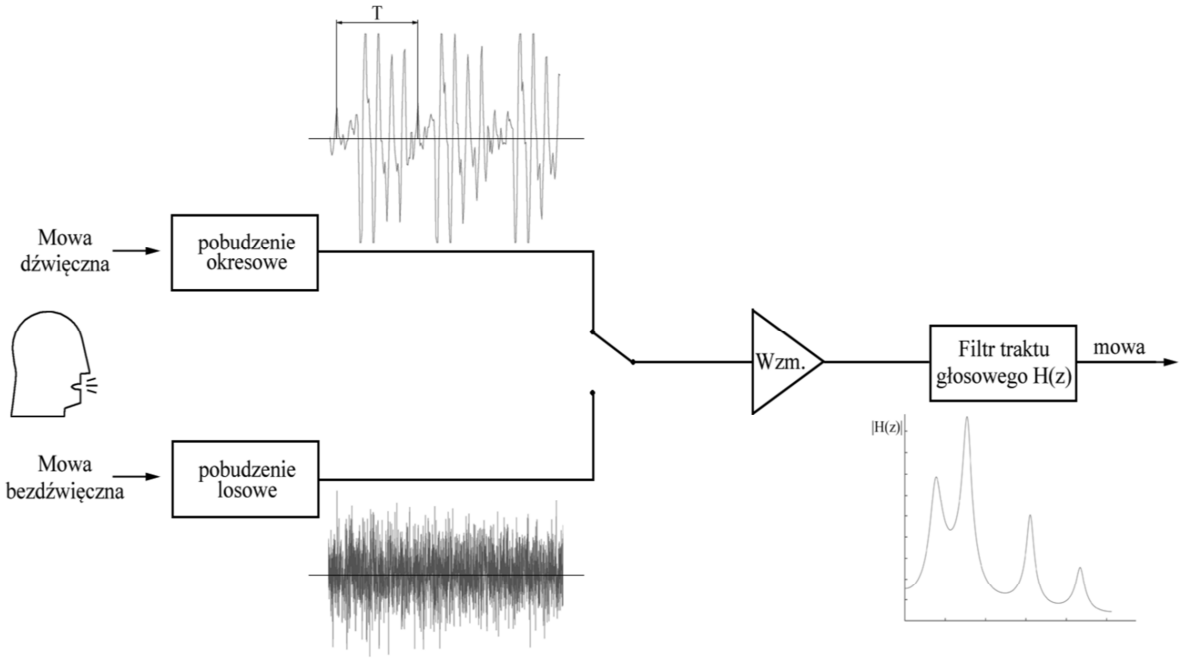
Artykulacja głosek bezdźwięcznych wymaga pobudzenia szumowego, które jest efektem szumu turbulencyjnego powstającego w wyniku przepływu laminarnego strumienia wydychanego powietrza przez wąską szczelinę w jamie ustnej (rys. 4.3). Artykulacja spółgłosek zwartych jest efektem fali udarowej powstającej w wyniku nagłego otworzeniu drogi przepływu powietrza



Rysunek 4.3 Model wytwarzania pobudzenia szumowego.

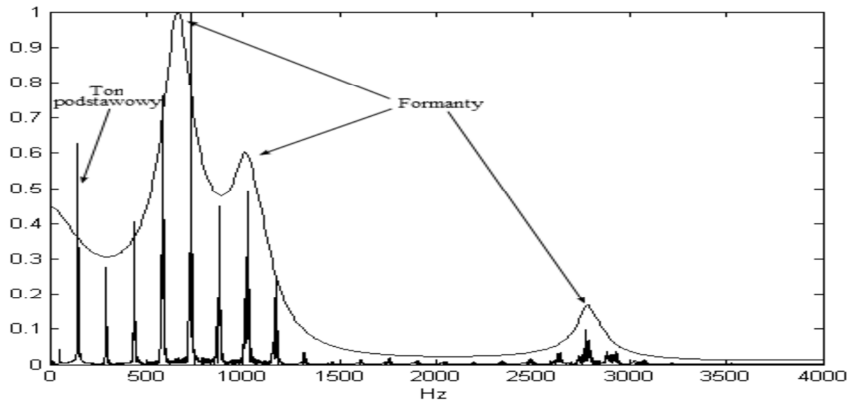
4.1.2. Proces powstawania sygnału mowy

Proces powstawanie sygnału mowy można przedstawić za pomocą modelu źródło-filtr, wykorzystywanego w kodowaniu, syntezie oraz rozpoznawaniu mowy. Model ten składa się z dwóch elementów źródła i filtru. Źródło modeluje sygnał pobudzający: dźwięczny lub bezdźwięczny (szumowy) w zależności od rodzaju wypowiadanej głoski. Charakterystyka amplitudowo-częstotliwościowa traktu głosowego (jama ustna, nosowa, język i usta) „formuje" sygnał pobudzający, w wyniku czego powstają charakterystyczne dla danej głoski lokalne maksima (formanty) decydujące o rozpoznawalności głoski (rys. 4.5). Zmiana ułożenia poszczególnych elementów traktu głosowego zmienia parametry filtru, co przekłada się na zmianę poziomów oraz położenie formantów, będących charakterystycznymi skupiskami energii wokół częstotliwości rezonansowych traktu. Wyznaczając charakterystykę częstotliwościową filtra traktu głosowego zgodnie z zależnościami wprowadzonymi w rozdziale 4.3 można określić lokalizacje formantów. Schemat obrazujący model wytwarzania mowy został pokazany na rys. 4.4.



Rysunek 4.4 Model wytwarzania sygnału mowy typu źródło-filtr.

Rozłożenie formantów na osi częstotliwości oraz ich względne poziomy są decydujące dla prawidłowego rozpoznawania głosek (rys. 4.7). Z uwagi na te własności są często wykorzystywane w procesach automatycznego rozpoznawaniu mowy (ARM). Ludzki słuch jest bardzo czuły na zmiany formantów, w związku z czym nawet niewielkie odchylenia częstotliwości formatowych mogą powodować dużą zmianę jakościową w odbiorze i zrozumiałości głosek. W skrajnych przypadkach zniekształceń liniowych powodujących zmiany poziomów i położenia formantów może dochodzić do przesunięcia formantów do obszarów charakterystycznych dla innych głosek.

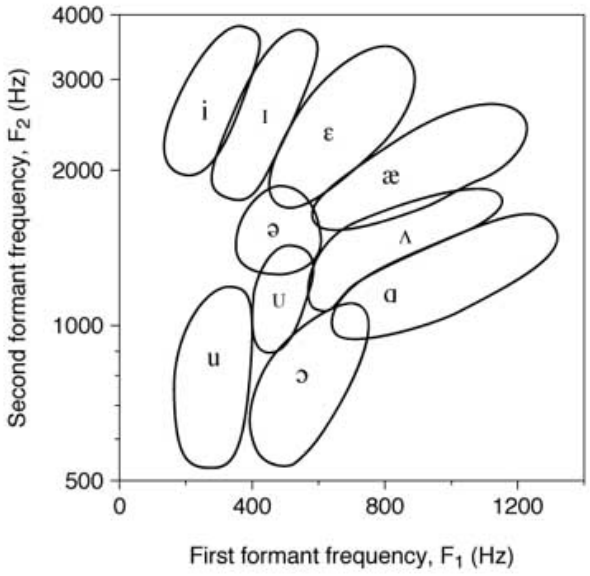


Rysunek 4.5 Obwiednia widma częstotliwościowego samogłoski „y” z zaznaczonymi formantami i tonem podstawowym.

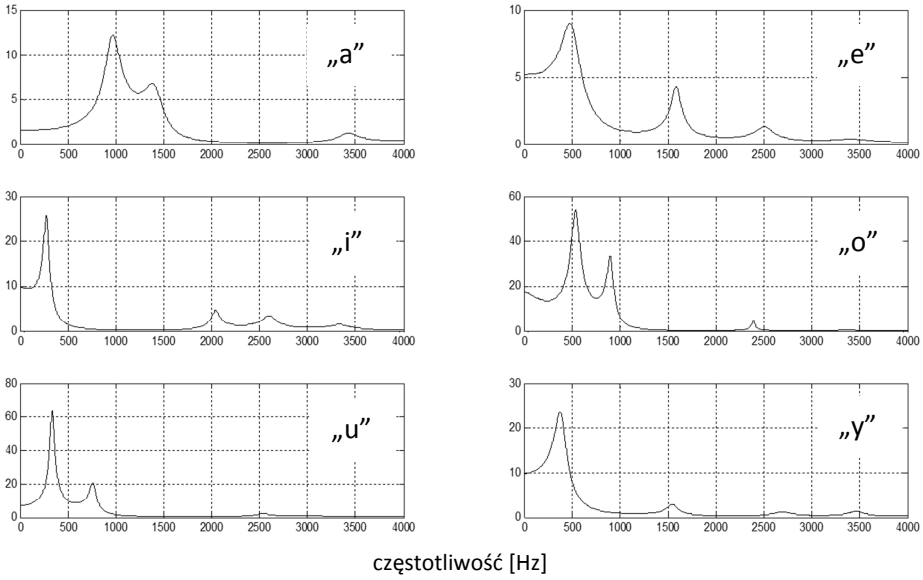
Tabela 4.2 Przykładowe wartości parametrów formantowych [28].

Fonem	Częstotliwości [Hz]				Poziomy względne [dB]			
	F1	F2	F3	F4				
i	210	2750	3500	4200	0	-15	-15	-27
e	380	2610	3000	3600	0	12	16	20
a	780	1150	2700	3500	0	-7	-25	-25
y	240	1550	2400	3300	0	-12	-20	-30
o	400	730	2300	3200	0	-3	-30	-35
u	270	615	2200	3150	0	-13	-40	-50
w	600	1700	2900	4100	-9	0	-2	-10
sz	-	2300	2900	3600	-	-9	-8	0
h	500	1700	2500	4200	-12	0	-10	-17
z	-	1750	2950	4300	-	-6	-10	0

W przypadku samogłosek najistotniejsze w procesie ich rozpoznawania są dwa pierwsze formanty: F1 i F2 (tabela 4.2). Wykres przedstawiający mapę obszarów względnego położenia dwóch pierwszych formantów, zwany „trójkątem samogłosek”, przedstawiono na rys. 4.6. Zakresy częstotliwości, w których zazwyczaj powstaje pierwszy bądź drugi formant dla różnych głosek mogą się pokrywać się (np. F1=280Hz może występować dla głosek „i”, „y” oraz „u”). W przypadku zniekształceń powodujących przesunięcia częstotliwości formantowych może więc dojść do sytuacji zmiany brzmienia jednej głoski w drugą.



Rysunek 4.6 Mapa rozłożenia formantów F1 i F2 dla samogłosek [29].



Rysunek 4.7 Przykładowe położenie formantów dla samogłosek, mowa męska.

Jak widać dla każdego fonemu, stan traktu głosowy mówcy daje się opisać zbiorem parametrów, które znajdują odzwierciedlenie w występowaniu formantów, czyli charakterystycznych maksimów widma chwilowego sygnału. Położenie oraz poziom formantów występujących w fonemach wpływa na możliwości ich rozróżnienia przez słuchacza. Oczywiście parametry formantowe nie opisują w pełni sygnału mowy, a są jedynie jego szczególną cechą pozwalającą na zobrazowanie niekorzystnych zjawisk zachodzących w złożonych układach akustycznych. Jak to wykazano w niniejszej pracy zniekształcenia sygnału wprowadzane przez kanał transmisyjny złożony z wielu szerokopasmowych źródeł bardzo silnie wpływają na zmianę poziomów i położenie poszczególnych formantów. Efekt ten pozwala więc na skuteczne wykorzystanie zmian parametrów formantowych do opisu zjawisk zachodzących w polu takich układów.

4.2. Parametryzacja sygnału mowy

Proces parametryzacji dowolnego obiektu polega na wyekstrahowaniu z niego wektora cech, który w formalny sposób opisuje ten obiekt. Ekstrakcja taka powinna charakteryzować się powtarzalnością być deterministyczna i posiadać formalizm matematyczny. Dzięki temu zamiast abstrakcyjnego opisu opartego na subiektywnych wrażeniach obserwatora obiektu, uzyskuje on opis pozwalający na obiektywne porównywanie jego cech w kategoriach liczbowych. Jest to szczególnie ważne w sytuacji porównywaniu cech obiektu z wykorzystaniem obliczeń komputerowych. Dzięki uściślonemu w wyniku parametryzacji formalnemu opisowi obiektów możliwe jest dokonanie klasyfikacji tych obiektów oraz obiektywne oszacowanie różnic pomiędzy nimi.

4.2.1. Parametryzacja w dziedzinie czasu

Podstawowym opisem sygnałów akustycznych jest ich prezentacja w dziedzinie czasu. Odpowiada ona zmianom poziomu ciśnienia akustycznego w czasie, którą opisać można za pomocą funkcji rzeczywistej *x(t)*. Biorąc pod uwagę model wytwarzania sygnału mowy, można ją przedstawić jako splot źródłowego pobudzenia *x(t)* i odpowiedzi impulsowej traktu głosowego *h(t)*.

W przypadku głosek dźwięcznych (wszystkie samogłoski oraz np. „z", „g", „w") pobudzeniem jest ton krtaniowy powstający w wyniku drgań strun głosowych, nazywany też formantem zerowym F_0. Dla głosek bezdźwięcznych pobudzenie ma charakter szumowy. Może to być szum turbulencyjny powstający w wyniku swobodnego wydmuchiwania strumienia powietrza (np. głoski „s", „f", „sz") lub fala udarowa powstająca po nagłym otworzeniu drogi przepływu powietrza (np. głoski „t", „p", „c").

Pierwotny opis sygnału jako przebiegu czasowego może zostać poddany przekształceniom matematycznym, w celu uzyskania zestawu parametrów. Na potrzeby przetwarzania cyfrowego sygnał mowy poddawany jest zazwyczaj procesowi próbkowania (czyli dyskretyzacji w dziedzinie czasu) oraz kwantyzacji wartości sygnału. Taka reprezentacja sygnału wymaga spełnienia twierdzenia Nyquista [25].

Dla sygnału mowy zobrazowanego zarówno w ciągłej jak i dyskretnej skali czasu można wyznaczyć szereg parametrów mających interpretację fizyczną ważną z punktu widzenia dalszej analizy stopnia zniekształcenia tego sygnału. Obliczenia wykonywane są dla odcinków czasowych określonych w procesie segmentacji sygnału.

Do podstawowych parametrów czasowych należą: wartość średnia sygnału, wartość minimalna i maksymalna sygnału, obwiednia amplitudy sygnału, funkcja gęstość przejść przez zero, wartość skuteczna sygnału, energia sygnału, moc średnia sygnału, środek ciężkości sygnału.

4.2.2. Parametryzacja w dziedzinie częstotliwości

Z biologicznego punktu widzenia, zarówno proces nadawania sygnału mowy (artykulacja), jak i jego odbiór, które związane są z procesami mózgowymi, polegają na modyfikowaniu oraz analizie widma sygnału. W procesie artykulacji następuje charakterystyczne formowanie obwiedni widma, natomiast proces słyszenia opiera się na rozdzielaniu składowych częstotliwościowych w sieci neuronowej mózgu. Opis sygnału za pomocą parametrów częstotliwościowych daje jednocześnie lepsze rezultaty, niż opis czasowy, w zakresie badań nad rodzajem i stopniem zniekształceń.

Analiza widmowa pozwala na wyodrębnienie składników amplitudowych i fazowych, które wpływają na formowanie się pola ciśnienia akustycznego. Daje także możliwość rozróżnienia źródła krtaniowego od elementów modulujących widmo w procesie artykulacji mowy.

Analizy spektralnej dokonuje się na ogół poprzez przekształcenie sygnału z dziedziny czasu na dziedzinę częstotliwości w oparciu o transformatę Fouriera. Dla sygnałów cyfrowych wykorzystywana jest dyskretna transformata Fouriera DFT. Do podstawowych parametrów widmowych należą:

- widmo amplitudowe sygnału (moduł DFT),
- funkcja widmowej gęstości mocy PSD (*ang. Power Spectral Density*),
- momenty widmowe (unormowane, centralne) *m*-tego rzędu, gdzie moment unormowany zerowego rzędu ma sens mocy sygnału, a moment unormowany pierwszego rzędu ma sens środka ciężkości widma (*ang. Spectral Centroid*),
- płaskość widmowa SFM (*ang. Spectral Flatness Measure*) – stosunek średniej geometrycznej i arytmetycznej współczynników widma – miara harmoniczności sygnału.

Podstawową procedurą wyjściową do ekstrakcji cech obiektów akustycznych, takich jak sygnał mowy, jest wyznaczenie krótkoczasowego widma sygnału STFT (*ang. Short-Time Fourier Transform*). Pozwala ono na przedstawienie zmienności sygnału mowy w czasie. Analiza taka stanowi kompromis pomiędzy rozdzielczością uzyskiwaną w dziedzinie czasu oraz w dziedzinie częstotliwości. Szerokie okno w dziedzinie czasu daje dużą rozdzielczość w dziedzinie częstotliwości ale kosztem małej rozdzielczości na osi czasu. Analogicznie wąskie okno czasowe daje niską rozdzielczość widmową. Nie jest możliwe uzyskanie wysokiej rozdzielczości jednocześnie na obu osiach. W zastosowaniach przetwarzania cyfrowego wykorzystuje się krótkoczasową transformatę dyskretną o postaci:

$$STFT(n,k) = \sum_{m=0}^{N-1} \gamma^*(m)x(n-m)e^{-j\left(\frac{2\pi}{N}k\right)m}, \quad k = 0,1,2,...,N-1 \tag{4.1}$$

gdzie:

$\gamma(m)$ - okno czasowe obserwacji

i wyznacza się ją przy pomocy algorytmu FFT szybkiej transformacji Fouriera [25].

4.2.3. Podejście perceptualne

W celu lepszego odwzorowania własności mechanizmu słyszenia ucha ludzkiego, charakteryzującego się nieliniową percepcją wysokości częstotliwości odbieranego dźwięku, wykorzystuje się model perceptualny. Prawo Webbera-Fechnera głosi, że „*Reakcja układu biologicznego jest proporcjonalna do logarytmu pobudzającego go bodźca*.” W ogólności można stwierdzić, że subiektywne wrażenie człowieka nie zależy w prosty sposób od obiektywnie mierzalnego pobudzenia. Oznacza to, że ludzkie ucho nie odpowiada liniowo na zwiększającą się częstotliwość. Aby dokonać analizy dźwięków w sposób uwzględniający to zjawisko, konieczne

jest zastosowanie przekształcenia skali częstotliwości z wykorzystaniem tzw. perceptualnych skali częstotliwości.

Najpowszechniejszą perceptualną skalą częstotliwości jest wykorzystywana w muzyce skala oktawowa. Odpowiada ona tzw. strojowi równomiernie temperowanemu. W badaniach na ludzkim słuchem, zrozumiałością oraz w innych zastosowaniach technicznych, takich jak kodowanie sygnału mowy, najpowszechniej wykorzystywane są skale barkowa i melowa.

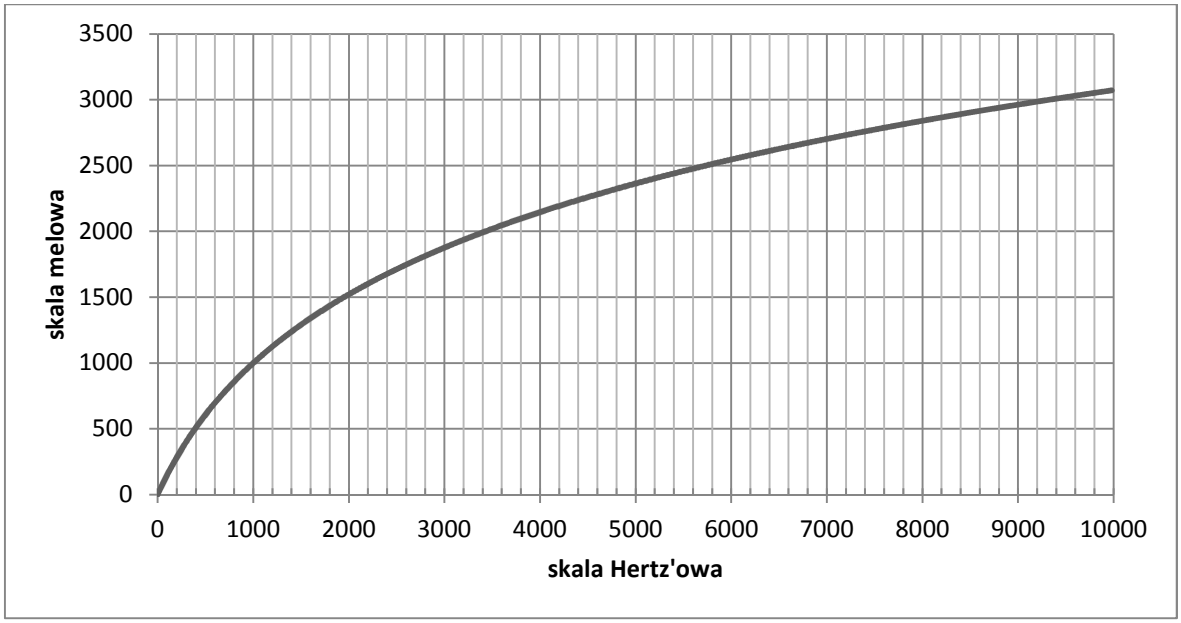
Skala barkowa wyznaczana jest w oparciu o szerokości pasm krytycznych, czyli takich zakresów częstotliwości, po przekroczeniu których odczuwana jest wyraźna zmiana głośności dźwięków. Wyróżnia się 24 pasma krytyczne.

Skala melowa wyznaczana jest w oparciu o tony proste. Odpowiada ona subiektywnemu wrażeniu wysokości dźwięku. W wyniku badań ustalono, że wrażenie wysokości dźwięku zależy również od głośności dźwięku, stąd w definicji skali przyjęto poziom natężenia 40dB SPL (względem 20μPa). Jako punkt odniesienia przyjmuje się częstotliwość 1 kHz, dla którego krzywa znajduje się 40 dB ponad progiem słyszenia człowieka i oznacza się go jako 1000 meli. Do wyznaczenia skali melowej (rys. 4.8) na podstawie skali częstotliwości stosuje się zależność:

$$f_{mel} = 2595 \log_{10}\left(1 + \frac{f_{Hz}}{700}\right) \tag{4.2}$$

Znajomość skali perceptualnych pozwala na wyznaczenie parametrów sygnałów dźwiękowych z uwzględnieniem naturalnej odpowiedzi układu słuchowego na pobudzenie dźwiękami mowy. Można np. wyznaczyć energię w pasmach melowych lub energię w pasmach krytycznych.

Najważniejsze z punktu widzenia metod badania zrozumiałości, rozpoznawania i kompresji mowy jest zastosowanie skali perceptualnych do wyznaczenia współczynników mel-cepstralnych MFCC (*ang. Mel-Frequency Cepstral Coefficients*), co zostało szerzej opisane w dalszej części niniejszego rozdziału.



Rysunek 4.8 Skala melowa wg. Beranka (1000 mel = 1000 Hz).

Analiza spektralna sygnału zawiera dużą ilość informacji, które są trudne do bezpośredniego zinterpretowania. Aby rozpoznać informacje istotne z punktu widzenia zmian zrozumiałości mowy konieczne jest wyselekcjonowanie takich cech widma sygnału, które determinują sposób jego percepcji przez odbiorcę.

Podstawowymi parametrami widma sygnału mowy związanymi z mechanizmem jego generowania są: ton krtaniowy F_0 (ton podstawowy) oraz formanty. Parametry formantowe charakteryzowane są poprzez częstotliwość formantu F_N oraz poziom formantu A_N, który może być mierzony jako bezwzględny lub względny, unormowany do najsilniejszego formantu i wyrażany w dB. Parametry formantów można wyznaczyć na podstawie obwiedni widmowej sygnału (*ang. spectral envelope*). Do najważniejszych metod wyznaczania widma wygładzonego należą parametry cepstralne oraz liniowe kodowanie predykcyjne.

4.3. Liniowa predykcja

Proces generowania mowy może być efektywnie modelowany za pomocą liniowej predykcji. W niniejszej rozprawie wykorzystano analizę sygnałów mowy opartą na liniowej predykcji do badań nad zniekształceniami sygnałów mowy w wielośródłowych układach akustycznych. Poniżej opisano sposób wyznaczanie współczynników predykcji oraz zastosowanie techniki LPC do modelowania charakterystyki traktu głosowego.

Krótkoczasowa (ang. *short-time*) korelacja obserwowana dla fragmentu sygnału mowy jest funkcją kształtu traktu głosowego. Sygnał mowy nie jest stacjonarny, parametry traktu głosowego zmieniają się w czasie. Długość fragmentu mowy, dla którego wyznaczane są parametry traktu jest

ograniczona. Przyjmuje się zazwyczaj, że w krótkim okresie 20-30 ms nie następują istotne zmiany tych parametrów.

Dla fragmentu sygnału mowy złożonego z N próbek {x[1], x[2], …, x[N]} wartość bieżącej próbki może zostać w przybliżeniu aproksymowana liniową kombinacją wartości p-poprzedzających ją próbek:

$$\hat{x}[n] = -\sum_{k=1}^{p} a_k x[n-k] \tag{4.3}$$

gdzie: p – jest rzędem predykcji, {a_1, …, a_p} – są to współczynniki predykcji, zwane dalej współczynnikami LPC. Przy kodowaniu sygnału spróbkowanego z częstotliwością 8 kHz typowo przyjmuje się rząd predykcji p=10 dla głosek dźwięcznych i p=4 dla głosek bezdźwięcznych (szumowych).

Błąd predykcji e_n pomiędzy rzeczywistą wartością próbki a jej aproksymowaną wartością wynosi:

$$e[n] = x[n] - \hat{x}[n] = x[n] - \sum_{k=1}^{p} a_k x[n-k] \tag{4.4}$$

Sygnał resztkowy *(ang. residual signal)* {e[n]} jest wyznaczany poprzez odjęcie sygnału estymowanego $\{\hat{x}[n]\}$ od sygnału oryginalnego $\{x[n]\}$. Krótkoczasowa korelacja pomiędzy próbkami sygnału resztkowego jest niska, a obwiednia jego widma mocy jest w przybliżeniu płaska.

Transformata z dla równana (4.4) wygląda następująco:

$$E(z) = A(z)S(z) \tag{4.5}$$

gdzie: $S(z)$ oraz $E(z)$ są to z-transformaty odpowiednio sygnału oryginalnego i sygnału resztkowego, natomiast:

$$A(z) = 1 + \sum_{k=1}^{p} a_k z^{-k} \tag{4.6}$$

Filtr $A(z)$ z równania (4.6) jest określany często mianem filtru „wybielającego”, który eliminuje korelację sygnału, a więc spłaszcza jego widmo.

Ponieważ $E(z)$ ma w przybliżeniu płaskie widmo, krótkoczasowa obwiednia widma mocy sygnału mowy jest modelowana autoregresywnie (AR) filtrem biegunowym:

$$H(z) = \frac{1}{A(z)} \tag{4.7}$$

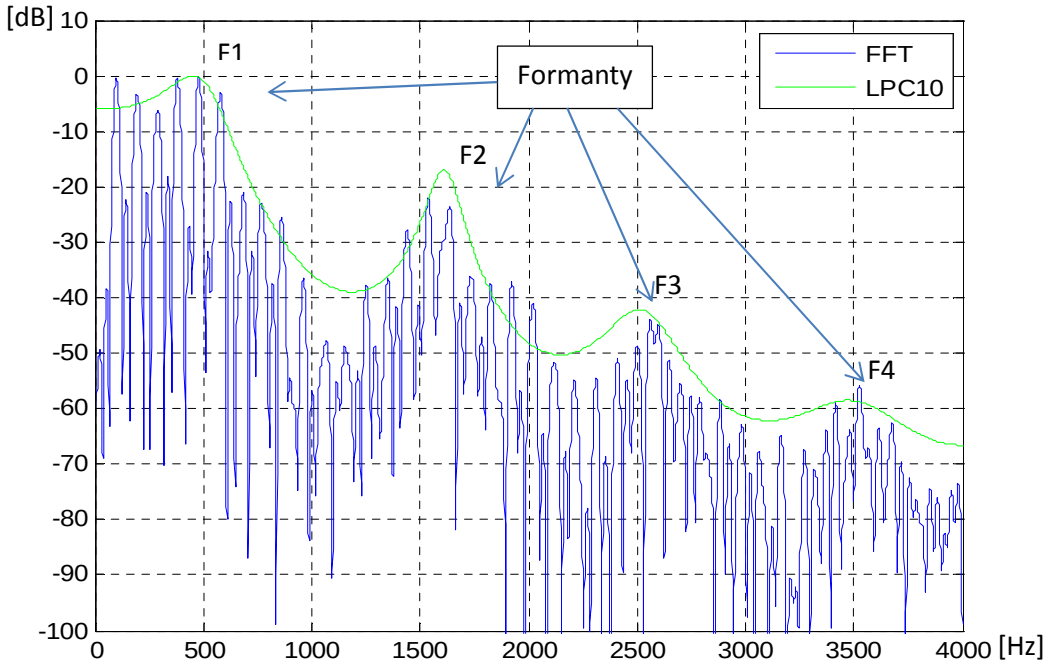
Filtr *A(z)* jest filtrem odwrotny do filtru sygnału mowy *H(z)*, jest to tzw. filtr „inwersyjny".

W analizie predykcyjnej krótkookresowa obwiednia widma mocy sygnału mowy jest otrzymywana poprzez wyznaczenie *H(z)* na okręgu jednostkowym.

Współczynniki predykcji są wyznaczane z sygnału mowy przy założeniu minimalizacji błędu średnio-kwadratowego:

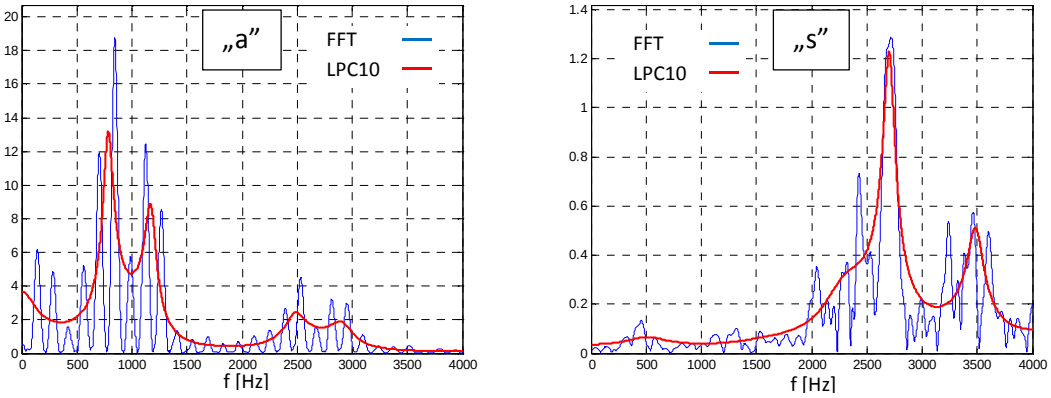
$$E = \sum_{n=-\infty}^{\infty} e[n]^2 \tag{4.8}$$

Liniowe kodowanie predykcyjne LPC (*ang. Linear Predictive Coding*) jest techniką kodowania sygnału mowy wykorzystującą predykcję liniową, tzn. polega ona na przedstawieniu sygnału mowy jako odpowiedzi filtru biegunowego AR (*ang. AutoRegresive*) na pobudzenie tonem krtaniowym. Kodowanie LPC odzwierciedla rezonansową charakterystykę traktu głosowego (rys. 4.9 - 4.11). Na podstawie znajomości parametrów filtru oraz pobudzenia można następnie odtworzyć sygnał pierwotny.

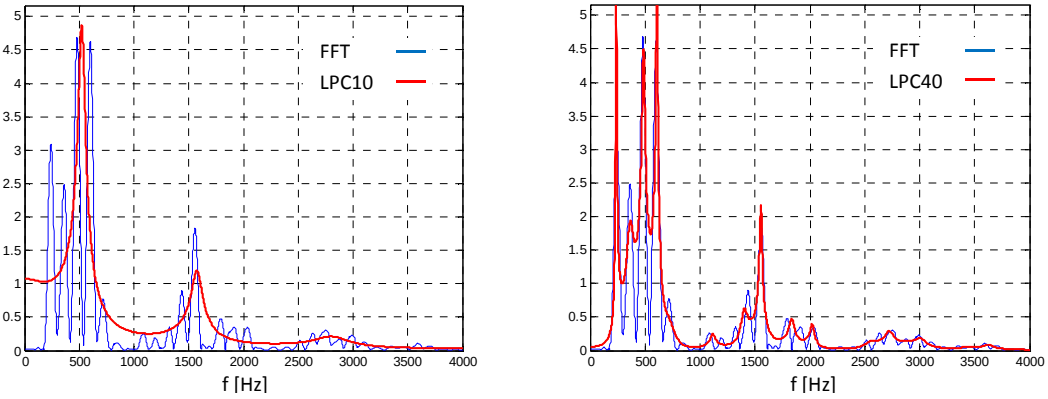


Rysunek 4.9 Widmo transmitancji filtru traktu głosowego dla głoski „e" z wyraźnymi formantami.

Na rysunkach 4.10 i 4.11 przedstawiono przykładowe kształty modułów filtru traktu głosowego przy zastosowaniu modelowania predykcyjnego. Jak widać na rys. 4.11 otrzymana w wyniku analizy LPC charakterystyka silnie zależy od doboru rzędu predykcji.



Rysunek 4.10 Przykłady widma transmitancji filtru traktu głosowego dla dwóch głosek, dźwięcznej „a” i bezdźwięcznej „s” nałożone na ich widmo FFT.



Rysunek 4.11 Przykład widma transmitancji filtru traktu głosowego dla głoski „e” nałożone na jej widmo FFT dla różnych rzędów predykcji 10 i 40.

4.3.1. Wyznaczanie współczynników predykcji metodą autokorelacji

Do wyznaczenia wartości współczynników predykcji można zastosować kilka alternatywnych metod takich jak: metoda autokorelacyjna lub metoda kowariancyjna [25]. Poniżej opisano metodę autokorelacyjną wykorzystywaną w pracy do obliczania współczynników predykcji.

Przy wyznaczaniu współczynników predykcji metodą autokorelacji, sumowanie wartości próbek sygnału powinno odbywać się w całym jego zakresie, tzn. w przedziale czasu $[-\infty,\infty]$. Jednak w przypadku analizy krótkoczasowej zdeterminowanej przez ograniczony czas stacjonarności sygnału mowy, operacja sumowania dokonuje się tylko dla fragmentu sygnału podlegającego analizie w danym momencie (w wyodrębnionej ramce sygnału). Przyjmuje się jednocześnie wartości zerowe dla wszystkich próbek sygnału spoza tego okna.

Do wyodrębnienia fragmentu sygnału mowy wykorzystywane są zazwyczaj opadające okna kosinusowe takie jak okno Hamminga lub Hanninga. Rzadziej stosowane jest okno prostokątne. Tak spreparowany sygnał poddawany jest następnie analizie predykcyjnej. W przypadku metody autokorelacyjnej zachowane jest kryterium minimalizacji błędu predykcji w postaci następującego równania:

$$\sum_{k=1}^{p} r_{|i-k|} a_k = -r_i, \ 1 \le i \le p \tag{4.9}$$

gdzie: r_k jest k-tym współczynnikiem autokorelacji dla analizowanego fragmentu mowy i wyznaczany jest zgodnie z następującym równaniem:

$$r_k = \frac{1}{N} \sum_{n=k}^{N} w_n x[n] w_{n-k} x[n-k] \tag{4.10}$$

gdzie: $\{w_i\}$ jest funkcją okna obejmującego N próbek sygnału.

Na podstawie (4.10) można wyznaczyć układ p równań znanych jako równania Yule-Walker'a, dzięki rozwiązaniu którego otrzymujemy p współczynników predykcji. Ten układ równań można zapisać w postaci macierzowej w formie następującego równania:

$$\mathbf{Ra}=\mathbf{-r} \tag{4.11}$$

gdzie:

$$R = \begin{bmatrix} r_0 & r_1 & r_2 & \cdots & r_{p-1} \\ r_1 & r_0 & r_1 & \cdots & r_{p-2} \\ r_2 & r_1 & r_0 & \cdots & r_{p-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_{p-1} & r_{p-2} & r_{p-3} & \cdots & r_0 \end{bmatrix} \tag{4.12}$$

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{p-1} \\ a_p \end{bmatrix} \tag{4.13}$$

$$r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_{p-1} \\ r_p \end{bmatrix} \tag{4.14}$$

Macierz ***R*** z równia (4.12) posiada strukturę Toeplitz'a i jest nazywana często macierzą autokorelacji.

Aby ułatwić wyznaczenie wektora $\{a_i\}$ współczynników predykcji, rozwiązania równań (4.11) dokonuje się w oparciu o szybkie algorytmy, jak na przykład algorytm Levinsona-Durbina [30] [25] [31] bądź algorytm Shur'a [32]. Posiadanie przez macierz ***R*** struktury Toeplitza gwarantuje stabilność filtru syntezy *H(z)* uzyskanego metodą autokorelacji, gdyż wszystkie jego bieguny będą się znajdowały zawsze wewnątrz okręgu jednostkowego. Jednocześnie w celu wzmocnienia słabych formantów w wyższych częstotliwościach, operację modelowania predykcyjnego dokonuje się zwykle na sygnale przefiltrowanym nierekursywnym filtrem FIR (preemfaza), transformującym sygnał oryginalny do postaci:

$$x'[n] = x[n] - 0.9375x[n-1] \tag{4.15}$$

W celu zdekodowania sygnału konieczne jest zastosowanie filtracji odwrotnej (deemfazy) przy zastosowaniu filtru rekursywnego IIR, transformującym zdekodowany sygnał do postaci:

$$x[n] = x'[n] + 0.9375x[n-1] \tag{4.16}$$

Na podkreślenie zasługuje fakt, iż w wielu przypadkach zamiast bezpośredniego stosowania współczynników predykcji wygodniej jest wykorzystywać inne parametry, które można wyznaczyć na podstawie współczynników filtru predykcyjnego. Do najczęściej stosowanych należą bieguny funkcji transmitancji *H(z)* oraz współczynniki odbicia (PARCOR).

Na podstawie współczynników odbicia można wyznaczyć kolejne współczynniki takie jak: parametry odwrotnej funkcji sinus ISP (*ang. Inverse Sine Parameters*) czy też współczynniki logarytmicznego stosunku przekrojów tuby akustycznej LAR (*ang. Log Area Ratio Parameters).*

Alternatywną metodą reprezentacji współczynników predykcji jest również metoda częstotliwości widma liniowego LSF (*ang. Line Spectral Frequencies*) znana również pod nazwą LPS (*ang. Line Spectrum Pair*). Współczynniki LSF są zerami dwóch wielomianów *P(z)* oraz *Q(z)* utworzonych na podstawie filtru inwersyjnego *A(z).* Wszystkie zespolone zera tych wielomianów znajdują się na okręgu jednostkowym, więc do ich opisu wystarczy tylko jeden parametr (częstotliwość lub kąt). Jej zaletą jest duża wydajność obliczeniowa.

4.4. Analiza cepstralna – współczynniki MFCC

Metoda wyznaczania współczynników MFCC jest obok predykcji liniowej podstawową formą parametryzacji sygnału mowy i polega na wyznaczeniu współczynników cepstralnych z uwzględnieniem podejścia perceptualnego. Polega ona na użyciu skali melowej do przekształcenia częstotliwości sygnału.

Analiza cepstralna opiera się na fakcie występowania okresowości w widmie sygnału. Cepstrum jest to transformata Fouriera z logarytmu widma i odpowiada ona dziedzinie czasu. Przy założeniu, że pobudzenie sygnału mowy (ton krtaniowy) jest w przybliżeniu ciągiem impulsów przesuniętych względem siebie o czas *T*, wówczas jego transformata Fouriera jest również sumą impulsów o pulsacji $\omega_0 = 2\pi / T$. W widmie pojawiają się więc wyraźne prążki, a ich kształt jest zależny od widma okna czasowego, za pomocą którego wycinany jest fragment sygnału. Można więc wyznaczyć odwrotną transformatę Fouriera z modułu widma aby wyznaczyć okres powtarzania się prążków widma, a co za tym idzie okres tonu pobudzającego. Chcąc usunąć modulację amplitudową widma pobudzenia należy przed obliczeniem transformaty odwrotnej dokonać operacji logarytmowania.

$$c(n) = \frac{1}{N}\sum_{k=0}^{N-1} \ln\left|\sum_{m=0}^{N-1} w(m)x(m)e^{-j2\pi km/N}\right| e^{\pm j2\pi kn} = \\ = F^{-1}\left(\ln\left|H\left(e^{j\Omega}\right)P\left(e^{j\Omega}\right)\right|\right) = F^{-1}\left(\ln\left|H\left(e^{j\Omega}\right)\right|\right) + F^{-1}\left(\ln\left|P\left(e^{j\Omega}\right)\right|\right) \tag{4.17}$$

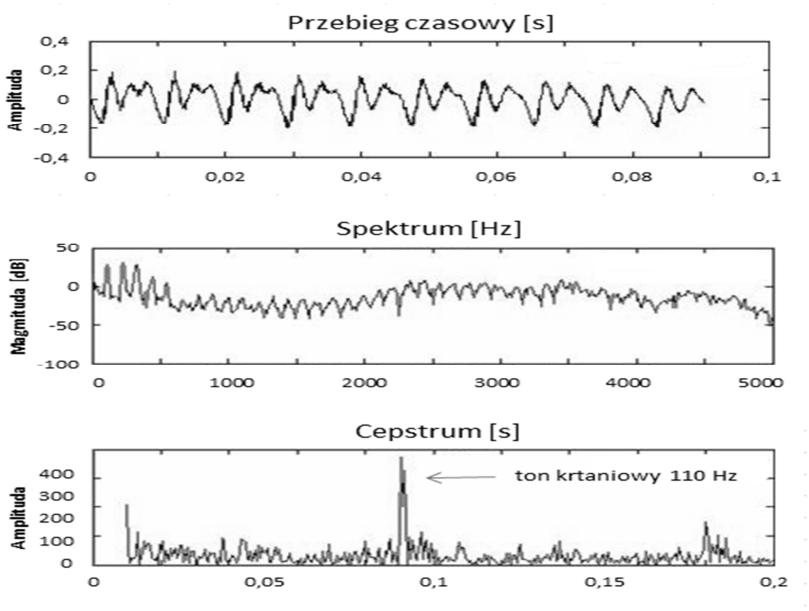
gdzie:

H(.) – widmo transmitancji traktu głosowego,

P(.) – widmo pobudzenia,

w(m) – okno czasowe (np. Hamminga).

W efekcie otrzymujemy dwie składowe w dziedzinie czasu odpowiadające wolnej zmienności obwiedni widma (wartości cepstrum w pobliżu indeksu zerowego) oraz prążki odpowiadające okresowi pobudzenia. Niskie współczynniki cepstralne niosą więc informację o charakterystyce traktu głosowego (wektor złożony z pierwszych kilkunastu współczynników) natomiast wysokie współczynniki cepstralne pozwalają na ekstrakcję tonu krtaniowego.



Rysunek 4.12 Wykres cepstrum z wyraźnym prążkiem tonu krtaniowego**.**

4.4.1. Wyznaczanie współczynników cepstralnych na podstawie LPC

Na podstawie znajomości współczynników liniowego kodowania predykcyjnego możliwe jest rekurencyjne wyznaczenie współczynników cepstralnych [30]. Możliwe jest również wyznaczenie ich w sposób bezpośredni [33]. Współczynniki cepstralne LPC oznaczone jako c_m można wyznaczyć z zależności:

$$\ln[H(z)] = \sum_{m=1}^{\infty} c_m z^{-m} \tag{4.18}$$

gdzie: *H(z)* - transmitancja filtru odwrotnego, opisana zależnością (4.7), w której:

$$A(z) = \sum_{k=0}^{M} a_k z^{-1} \tag{4.19}$$

gdzie a_0=1, M=2K+1,

K – liczba formantów, jaka jest brana pod uwagę w modelu.

Różniczkując (4.19) otrzymujemy zależność rekurencyjną pozwalającą wyznaczyć kolejne współczynniki cepstralne c_m:

$$c_m = -a_m - \frac{1}{m}\sum_{k=1}^{m-1} k c_k a_{m-k}, \; dla \; m > 0, \tag{4.20}$$

gdzie: a_0=1, a_k=0 dla k >M.

Ważone współczynniki cepstralne za pomocą wag w_m mają postać:

$$cw_m = c_m w_m, \quad 1 \le m \le M \tag{4.21}$$

gdzie:

$$w_m = 1 + \tfrac{M}{2} \sin\left(\frac{\pi m}{M}\right), \quad 1 \le m \le M \tag{4.22}$$

W metodach badawczych przyjmuje się zazwyczaj M=12. Współczynniki cepstralne są często wykorzystywane w procesie rozpoznawania mowy.

Jedną z miar wykorzystującą współczynniki cepstralne jest odległość cepstralna LPC CD (*ang. LPC cepstrum distance*), która jest definiowana jako [34]:

$$d_{LPCCD}{}^2(i) = \sum_{m=1}^{M} \left[c_m^c(i) - c_m^d(i)\right]^2 \tag{4.23}$$

gdzie:

i – numer ramki,

$c_m^c(i), c_m^d$ - współczynniki cepstralne odpowiednio oryginalnego i zniekształconego sygnału.

4.4.2. Współczynniki mel-cepstralne MFCC

Ludzkie ucho dokonuje rozróżnienia częstotliwości dźwięków w oparciu o nieliniową skalę widma sygnału. Skala ta jest w przybliżeniu liniowa tylko do częstotliwości ok. 1 kHz. Powyżej tej częstotliwości występuje nieliniowość, którą można przybliżyć odwzorowaniem logarytmicznym. Aby uwzględnić to zjawisko wykorzystywane są różne skale perceptualne. Najbardziej popularne są skala barkowa i melowa (rys. 4.8).

Wyznaczenie współczynników sygnału mowy uwzględniających subiektywny odbiór częstotliwości ludzkiego narządu słuchu, wymaga przekształcenia skali częstotliwości. Przekształcenie to dokonywane jest najczęściej z użyciem banku filtrów o częstotliwościach środkowych rozmieszczonych równomiernie na nieliniowej skali perceptualnej, co odpowiada nieliniowemu rozmieszczeniu na skali częstotliwości.

Do wyznaczania banku filtrów trójkątnych wykorzystuje się zależność (4.2), dzięki której trójkątne funkcje wagowe stają się niesymetryczne. Funkcje te o podstawie rzędu 200 lub 300 melów są nakładane na transformowany sygnał z przesunięciem 100 lub 150 melów.

Wyznaczenie współczynników cepstralnych z uwzględnieniem podejścia perceptualnego polega więc na użyciu skali melowej do przekształcenia częstotliwości sygnału [35]. Algorytm wyznaczania współczynników mel-cepstralnych MFCC obejmuje:

- Poddanie oryginalnego sygnału mowy operacji preemfazy zgodnie z (4.15).
- Wymnożenie sygnału oknem Hamminga.
- Obliczenie szybkiej transformaty Fouriera FFT dla poszczególnych segmentów sygnału mowy oraz wyznaczenie modułu estymaty widmowej gęstości mocy sygnału.
- Wykonanie filtracji melowej za pomocą zestawu 12 środkowoprzepustowych filtrów trójkątnych o częstotliwościach wyznaczonych zgodnie z (4.2).
- Zlogarytmowanie 12 uśrednionych wartości estymaty widma gęstości mocy oraz wyznaczenie ich transformat odwrotnych IFFT. Ponieważ widmo jest rzeczywiste i symetryczne, odwrotna transformata Fouriera redukuje się do dyskretnej transformaty kosinusowej obliczanej zgonie ze wzorem:

$$MFCC_n = \sqrt{\frac{2}{N}} \sum_{i=1}^{N} \log(S_i) \cdot \cos\left[\frac{\pi n}{N}(i - 0.5)\right] \tag{4.24}$$

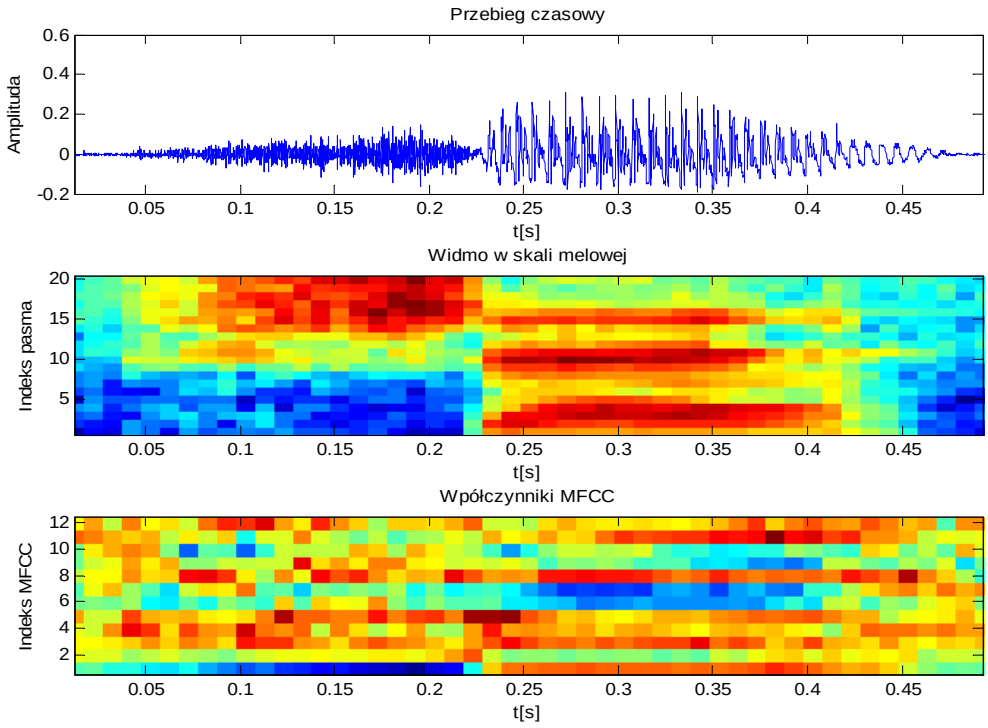
gdzie:

$MFCC_n$ – n-ty współczynnik mel-cepstralny,

S_i - uśredniona wartość estymaty widma gęstości mocy przy użyciu i-tego filtru,

N – liczba filtrów melowych , N=12.

Zamiast skali melowej można wykorzystać również skalę barkową.



Rysunek 4.13 Przebieg czasowy słowa (logatomu) „wsze” (góra), widmo w skali melowej (środek), współczynniki MFCC w kolejnych ramkach (dół).

4.5. Podsumowanie

Proces oceny jakości sygnału mowy (podobnie jak i proces rozpoznawania mowy) jest związany z własnościami perceptualnymi słuchu ludzkiego. Metoda wyznaczania współczynników MFCC jest, obok predykcji liniowej, podstawową formą parametryzacji sygnału mowy, wykorzystywaną głównie w systemach rozpoznawania mowy. Warto zauważyć, iż istnieją również techniki wykorzystujące współczynniki predykcji liniowej uwzględniające podejście perceptualne jak choćby PLP (*ang. Perceptual Linear Prediction*) czy też będąca jej rozwinięciem technika RASTA (*ang. RelAtive SpecTraA*) [36] [37].

Jakość pola akustycznego można opisać poprzez zmiany wartości parametrów przestrzenno-czasowych sygnału mowy w funkcji ilości oraz rozmieszczenia źródeł dźwięku. Polem akustycznym o dobrej jakości jest wówczas takie, w którym nie następują zmiany jego parametrów bądź są one stosunkowo niewielkie. Badania przeprowadzone na charakterystycznych cechach sygnałów mowy pozwalają ocenić wpływ tych różnic na jakość sygnałów mowy transmitowanych w systemie nagłośnienia. Rozłożenie formantów na osi częstotliwości oraz ich względne poziomy są decydujące dla prawidłowego rozpoznawania głosek. Ludzki słuch jest bardzo czuły na zmiany formantów, w związku z czym nawet niewielkie odchylenia częstotliwości formatowych mogą powodować dużą zmianę jakościową w odbiorze i zrozumiałości głosek. Jednocześnie formanty szczególnie silnie ulegają zniekształceniom w obszarze występowania interferencji szerokopasmowej. Do ekstrakcji parametrów formantowych wykorzystano analizę predykcyjną, a zastosowanie odpowiednich miar odległości pomiędzy wektorami wartości współczynników LPC daje możliwość ilościowej oceny zniekształceń występujących w polu akustycznym układów wielоźródłowych.

5. METODY BADANIA JAKOŚCI I ZROZUMIAŁOŚCI SYGNAŁU MOWY

Na jakość i zrozumiałość sygnału mowy wpływ ma wiele czynników takich jak: zakłócenia szumowe, błędy w artykulacji, odbicia i rewerberacje w pomieszczeniach, jakość urządzeń toru fonicznego, a także rozmieszczenie przestrzenne źródeł dźwięku. W bezpośrednich pomiarach nie zawsze można odseparować wpływ poszczególnych czynników. Rodzaj zniekształceń jakim poddawany jest sygnał pierwotny determinuje sposób w jaki badane będą zmiany tego sygnału oraz wpływ tych zmian na jakość i zrozumiałość przekazu. Techniki badania oraz wybór odpowiednich parametrów zależą więc od tego w jaki sposób powstają zniekształcenia i co jest ich główną przyczyną. Inny charakter będą miały zmiany sygnału w wyniku kodowania sygnału, inne w wyniku błędów transmisji w kanale telefonicznym, a jeszcze inne w pakietowej transmisji głosu.

Istnieją dwie główne kategorie metod pomiaru jakości sygnału mowy przekazywanego przez łańcuch komunikacyjny. Pierwsza, to metody subiektywne, w których grupa słuchaczy testuje zrozumiałość ilościowo (miarą zniekształceń jest procent błędnie odczytanych logatomów, wyrazów czy zdań) bądź jakościowo, wyrażając ogólną opinię o jakości przekazu w szkolnej skali ocen; wyniki zostają uśrednione, a metoda jest nazywana MOS (*ang. Mean Opinion Score*). Druga kategoria, to metody obiektywne, w których ocena dokonywana jest na podstawie pomiarów parametrów systemu niezależnych od słuchacza.

Z uwagi na dużą czasochłonność, kosztowność oraz brak pełnej powtarzalności odsłuchowych testów „subiektywnych", które odnoszą się bezpośrednio do istoty komunikacji, ciągle doskonalone są metody „obiektywne", których wyniki pomiarowe są powtarzalne dzięki wyeliminowaniu czynnika ludzkiego. Chodzi tu o analizę samego sygnału i stopnia jego zniekształcenia po przejściu przez system komunikacji. Jednakże wyzwaniem pozostaje opracowanie takiej metody analizy parametrów odbieranego sygnału mowy, która pozwoliłaby zdefiniować obiektywne miary zniekształceń mowy, pozostające w dobrej korelacji z subiektywnymi ocenami zrozumiałości.

Aby dokonać prawidłowej oceny jakości przekazu sygnału należy zastosować metodę pomiaru adekwatną do rodzaju badanych sygnałów oraz dostosowaną do charakteru zniekształceń. Metoda badania implikuje wybór niezbędnych parametrów opisujących parametry sygnału i kanału komunikacyjnego, a także zastosowane wskaźniki oraz miary różnic pomiędzy wartościami wskaźników.

5.1. Miary odległości pomiędzy wektorami parametrów

Porównywanie wektorów zawierających parametry sygnałów dźwiękowych mowy oryginalnej (czystej, bez zniekształceń) i zniekształconej wymaga zastosowania odpowiednich miar odległości. Parametryzację sygnałów oryginalnego oraz zniekształconego można uzyskać w wyniku takich analiz jak np.: kodowanie z liniową predykcją, analiza cepstralna, metoda częstotliwości widma

liniowego LFS bądź percepcyjna analiza predykcyjna PLP (*ang. Perceptual Linear Predictive*). Określenie odległości *d(a, b)* pomiędzy otrzymanymi wektorami parametrów *a* i *b*, pozwala oszacować podobieństwo, a co za tym idzie również i stopień zniekształcenia, porównywanych fragmentów dźwięku.

5.1.1. Własności miar odległości

Dobra miara odległości wykorzystywana do weryfikacji podobieństwa krótkich fragmentów dźwięków powinna posiadać następujące własności matematyczne:

- powinna być symetryczna, tzn.:

$$d(a,b) = d(a,b) \tag{5.1}$$

  gdzie: *a, b* – wektory parametrów wyekstrahowanych z sygnałów odpowiednio oryginalnego i zniekształconego,

- powinna być nieujemna, tzn.:

$$d(a,b) \geq 0 \tag{5.2}$$

- powinna spełniać równanie trójkąta, tzn.:

$$d(a,b) < d(a,c) + d(b,c) \tag{5.3}$$

- powinna uwzględniać perceptualne zróżnicowanie odbioru składowych częstotliwościowych w widmie mocy sygnału, tzn. jeżeli energia sygnału $s(f_1)$ zgromadzona w paśmie wokół częstotliwości f_1 jest percepcyjnie istotniejsza od energii $s(f_2)$ zgromadzonej w paśmie wokół częstotliwości f_2, wówczas:

$$d[a, a + \Delta s(f_1)] > d[a, a + \Delta s(f_2)] \tag{5.4}$$

5.1.2. Stosowane miary odległości

Obiektywne miary odległości operujące na parametrach sygnałów dźwiękowych są wynikiem obliczenia odległości pomiędzy dwoma wektorami złożonymi z obiektywnych parametrów sygnału takich jak np. współczynniki LPC. Obliczenia te wykonuje się zazwyczaj w oparciu o dobrze znane miary takie jak miara Euklidesowa czy Mehalanobisa.

Odległość Euklidesowa pomiędzy dwoma wektorami współczynników jest definiowana następująco:

$$d(a,b) = \sqrt{\sum_{i=1}^{P} \left(a[i] - b[i]\right)^2} \tag{5.5}$$

gdzie: $i = \{1, \ldots, P\}$ – indeks parametrów sygnału.

Jest to najbardziej typowy sposób badania odległości pomiędzy wektorami. Nie uwzględnia on jednak właściwości perceptualnych narządu słuch i z tego powodu nie najlepiej odwzorowuje zmiany związane z parametryzacją mowy.

Inne często stosowane miary odległości to:

- Miara Czybyszewa:

$$d_C(a,b) = \max_{i=1,..,P} |a[i] - b[i]| \tag{5.6}$$

- Miara Minkowskiego:

$$d_M(a,b) = \sum_{i=1}^{P} |a[i] - b[i]| \tag{5.7}$$

- Uogólnieniem miar standardowych jest odległość Mehalanobisa, którą można zapisać w postaci:

$$d(a,b)^2 = \sum_{i=1}^{P} \left[\frac{a[i] - b[i]}{\sigma_i} \right]^2 \tag{5.8}$$

gdzie: σ_i^2 – jest odchyleniem standardowym dla *i*-tego współczynnika.

- Miara odległości Log-Spectral:

$$d_{\log-S}(a,b) = \sqrt[q]{\sum_{i=1}^{P} |\log(a[i]) - \log(b[i])|^q} \tag{5.9}$$

gdzie: q=1,2, …, ∞ oraz a, $b > 0$.

- Miara odległości Itakura-Saito (IS):

$$d_{IS}(a,b) = \sum_{i=1}^{P} \left(\frac{a[i]}{b[i]} - \log \frac{b[i]}{a[i]} - 1 \right) \tag{5.10}$$

Miara IS jest niesymetryczna ale jest często stosowana z uwagi na dużo lepsze, niż w przypadku metryk Euklidesowych, odwzorowywanie własności psychoakustycznych, związanych z nieliniową percepcją zmian częstotliwości przez słuch ludzki. Problem braku symetrii miary IS można rozwiązać np. poprzez symetryzację do tzw. metryki „Cos-h” w postaci:

$$d_{Cosh}(a,b) = \frac{d_{IS}(a,b) + d_{IS}(a,b)}{2} \tag{5.11}$$

Metryki IS oraz Cos-h, pierwotnie zdefiniowane do oceny zmian współczynników LPC, mogą być z powodzeniem stosowane do innych parametrów związanych z częstotliwościową reprezentacją sygnału (np. MFCC).

5.1.3. Miary zniekształceń oparte na stosunku sygnału do szumu

Stosunek sygnału do szumu SNR (*ang. Signal to Noise Ratio*) jest jedną z najbardziej klasycznych miar zakłóceń sygnału. Jako miara intruzyjna wykorzystuje do określenia stopnia zniekształcenia wartości sygnału oryginalnego i zniekształconego. Stosunek Sygnału do szumu może być obliczany zarówno w dziedzinie czasu jak i dziedzinie częstotliwości. Jego postać czasową można zdefiniować następująco:

$$SNR = 10\log\frac{\sum_{n=1}^{N} x^2[n]}{\sum_{n=1}^{N}\left(x[n]-y[n]\right)^2} \tag{5.12}$$

gdzie: $x[n]$ to próbki sygnału wzorcowego (niezakłóconego), $y[n]$ to próbki sygnału zakłóconego, N – całkowita liczba próbek w badanym fragmencie sygnału.

Pomiar stosunku sygnału do szumu, w przypadku zakłóceń o innym charakterze niż tylko addytywne zakłócenia szumowe, może nie dawać dostatecznie dokładnych wyników. Dlatego powstało kilka odmian klasycznego SNR, które są dużo bardziej skorelowane z jakością mowy mierzoną w sposób subiektywny.

Klasyczny SNR uśrednia stosunek sygnału do szumu w ramach całkowitego czasu trwania sygnału, podczas gdy sygnał mowy jest sygnałem niestacjonarnym. Energia sygnału mowy fluktuuje w czasie, dlatego stosunek sygnału do szumu również powinien zmieniać się w czasie. Segmentowy stosunek sygnału do szumu SNR_{seg} (*ang. time-domain segmental SNR)*, jest obliczany jako średni SNR ze stosunku sygnału do szumu liczony w kolejnych fragmentach, na jakie podzielony jest cały przebieg badanego sygnał mowy:

$$SNR_{seg} = \frac{10}{M}\sum_{m=0}^{M-1}\log\frac{\sum_{n=Lm}^{Lm+L-1} x^2[n]}{\sum_{n=Lm}^{Lm+L-1}\left(x[n]-y[n]\right)^2} \tag{5.13}$$

gdzie L jest liczbą próbek w jednym fragmencie sygnału, M jest liczbą fragmentów w całym analizowanym przebiegu sygnału ($N=ML$). Czas trwania segmentu wynosi zwykle od 10 do 30 ms.

Kolejną odmianą stosunku sygnału do szumu jest ważony częstotliwościowo SNR. Ważenie $\mathrm{fwSNR_{seg}}$ (*ang. frequency weighted SNR*) następuje w podpasmach częstotliwości proporcjonalnych do pasma krytycznego.

$$\mathrm{fwSNR_{seg}} = \frac{10}{\mathrm{M}} \sum_{\mathrm{m}=0}^{\mathrm{M}-1} \frac{\sum_{\mathrm{j}=0}^{\mathrm{K}-1} \mathrm{W(j,m)} \log \frac{\mathrm{X(j,m)}^2}{\{\mathrm{X(j,m)} - \mathrm{Y(j,m)}\}^2}}{\sum_{\mathrm{j}=0}^{\mathrm{K}-1} \mathrm{W(j,m)}} \tag{5.14}$$

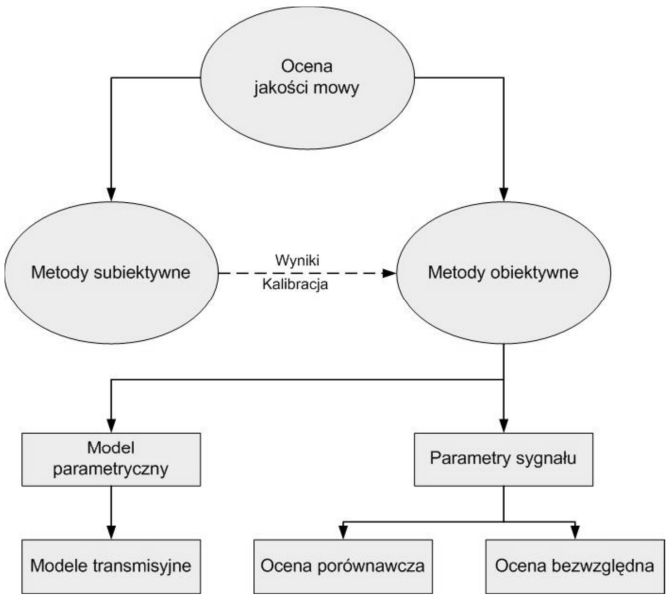
gdzie $\mathrm{W(j,m)}$ jest wagą j-tego podpasma w m-tym segmencie, K jest liczbą podpasm, $\mathrm{X(j,m)}$ jest widmem amplitudowym w j-tym podpaśmie m-tego segmentu, natomiast $\mathrm{Y(j,m)}$ jest widmem segmentu sygnału zakłóconego w tym samym podpaśmie. Wykorzystywane są dwa sposoby wagowania: stałe wagi związane z własnościami psychoakustycznymi lub zmienne wagi związane z mechanizmem generowania mowy. Jako zbiór wag $\mathrm{W(j,m)}$ można wykorzystać wagi określone w standardzie ANSI S3.5-1969 [38].

5.2. Metody badania jakości i zrozumiałości przekazu głosowego

Miarą zadowolenia odbiorcy z przekazu głosowego jest jego jakość, a właściwie subiektywne odczucie tej jakości przez każdego ze słuchaczy. Jednak zróżnicowanie cech artykulacyjnych osób generujących przekaz głosowy oraz percepcji słuchowej odbiorców powodują, że wrażenie jakości są mocno zróżnicowane u poszczególnych osób. Wrażenie to można oszacować poprzez wyciągnięcie średniej wartość z wielu subiektywnych ocen jakości dźwięku MOS zapisanych w określonej skali np. od 1 do 5. Miara MOS jest obecnie traktowana jako referencyjna miara odniesienia zarówno dla metod obiektywnych jak i metod subiektywnych pomiaru jakości sygnału mowy (tabela 5.1).

Badania oparte na subiektywnym wrażeniu jakości u słuchaczy wymaga przeprowadzenia wielu porównywalnych, przeprowadzonych w bardzo sformalizowany sposób odsłuchów, co jest bardzo trudne organizacyjnie, wymaga dużych nakładów czasu i wiąże się z wysokimi kosztami związanymi z zaangażowaniem wykwalifikowanej grupy lektorów i słuchaczy oraz powtarzania testów dla każdego wariantu badanego systemu.

Biorąc to pod uwagę opracowano szereg metod automatycznego szacowania jakości przekazu, które skorelowane są z wynikami ocen subiektywnych. Takie badania, nazywane często pomiarami obiektywnymi, opierają się głównie na porównywaniu parametrów odpowiednio dobranego sygnału oryginalnego o wysokiej jakości i sygnału zniekształconego, który dociera do słuchacza. Następnie na podstawie obliczeń wyznacza się wskaźniki odpowiadające skali MOS.



Rysunek 5.1 Klasyfikacja metod badania jakości mowy [39].

Oprócz obiektywnych porównawczych metod badawczych wykorzystujących sygnał wzorcowy, tzw. „metody intruzyjne” (*ang. intrusive*), istnieją inne, trudniejsze do zaimplementowania, tzw. „metody nieintruzyjne” (*ang. nonintrusive*), które nie wymagają znajomości sygnału wzorcowego. Metody obiektywne posługują się modelem percepcji aby ocenić jakość przekazu postrzeganą przez odbiorcę. Jeszcze inną grupę stanowią metody parametryczne wykorzystujące do oceny jakości jedynie parametry systemu komunikacyjnego, niezależne od rodzaju sygnału.

Spośród wielu metod badania jakości w pełni akceptowalne są jedynie metody subiektywne, które wymagają jednak spełnienia wielu trudnych wymogów takich jak: odpowiednio duża liczba słuchaczy, wielokrotność powtórzeń odsłuchu, odpowiednie odstępy w czasie pomiędzy poszczególnymi testami, odpowiednie warunki w pomieszczeniu odsłuchowym, zmęczenie słuchaczy i czynniki indywidualne wpływające na wyniki, konieczność przeprowadzania analizy statystycznej dla uzyskania porównywalności ocen itp. Badania subiektywne są więc drogie i czasochłonne. W praktycznych zastosowaniach, dużo wydajniejsze są metody obiektywne dające szybkie i powtarzalne wyniki. Aby uzyskać dużą korelację z wynikami pomiarów subiektywnych, co stanowi jeden z głównych warunków akceptowalności metod obiektywnych, konieczna jest ich odpowiednia kalibracja.

Tabela 5.1 Współczynniki korelacji metod obiektywnych z MOS [39].

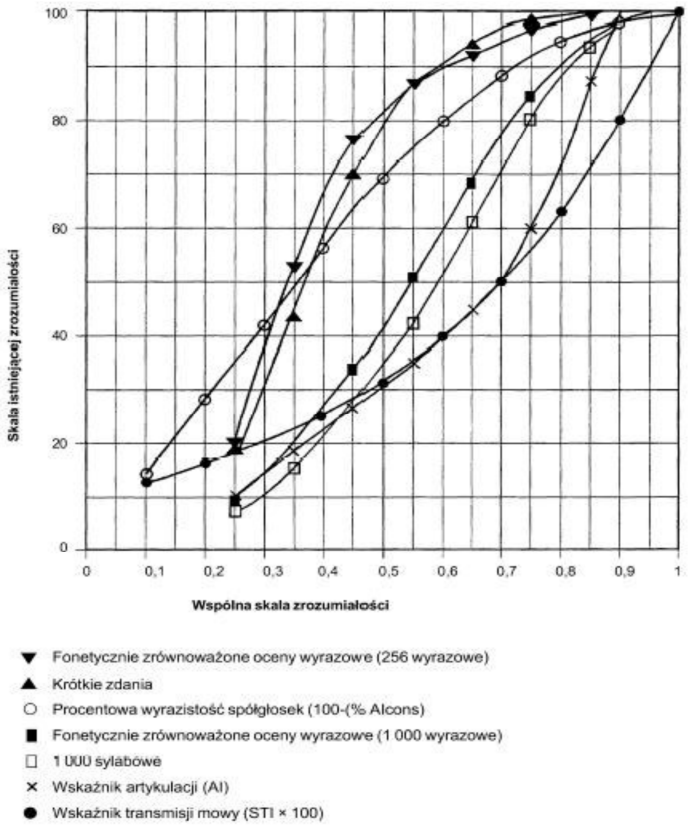
Typ sieci	Współczynnik korelacji	PSQM	PSQM+	PAMS	PESQ
GSM	Wartość średnia	0,924	0,935	0,954	0,962
	Najgorszy przypadek	0,843	0,859	0,895	0,905
PSTN	Wartość średnia	0,881	0,897	0,936	0,942
	Najgorszy przypadek	0,657	0,652	0,805	0,902
VoIP	Wartość średnia	0,674	0,726	0,916	0,918
	Najgorszy przypadek	0,260	0,469	0,758	0,810

Do testów jakości wykorzystywany jest materiał słowny, który składa się na dwa podstawowe rodzaje badań: zrozumiałości zdań i zrozumiałości słów. Różnica między rozpoznawaniem całego zdania, a identyfikacją pojedynczych słów jest zwykle związana z kontekstem. Codzienne wypowiedzi zawierają wiele informacji kontekstowych, co pomaga słuchaczowi rozpoznać niezrozumiałe słowa w przedstawionym zdaniu.

Nie można jednoznacznie stwierdzić czy kontekst wypowiedzi zwiększa, czy zmniejsza stosowalność i rzetelność badań. Z jednej strony, badania mają odzwierciedlać naturalny proces komunikacji, którego częścią jest kontekst. Z drugiej zaś wywołuje to efekt pamięciowy, który jest niepożądany w tego typu badaniach. Testy zdaniowe można więc podzielić na takie, które wzorowane są na codziennych wypowiedziach oraz te z użyciem semantycznie nieprzewidywalnych ciągów wyrazów tworzących nie zawsze logiczną wypowiedź.

Wspólna skala zrozumiałości CIS (ang. Common Intelligibility Scale)

Aby porównać wyniki pomiarów zrozumiałości przeprowadzonych różnymi metodami konieczne jest przeliczenie ich na wspólną skalę. W zaleceniu IEC 60849 [40] [8] zamieszczono wykres (rys. 5.2) zawierający graficzne porównanie pomiędzy wynikami pomiarów różnymi metodami oceny zrozumiałości. Zalecenie to, dotyczące dźwiękowych systemów ostrzegawczych DSO, wymaga aby zrozumiałość określona względem wspólnej skali zrozumiałości CIS wynosiła minimum 0.7. Dzięki wspólnej skali można określić minimalną wartość zastosowanego wskaźnika odpowiadającą temu kryterium.



Rysunek 5.2 Wspólna skala zrozumiałości CIS na podstawie PN-EN 60849 [8].

5.3. Metody subiektywnej oceny jakości przekazu głosowego

Metody subiektywne oceny jakości przekazu głosowego są stosowane jako podstawowy mechanizm kalibracji innych metod badania jakości przekazu mowy. Opierają się one na bezpośrednim odsłuchu przekazu przez grupę słuchaczy. W ten sposób oceniane są cechy sygnału mowy, które pozwalają na oszacowanie jego jakości. Subiektywne pomiary jakości mowy przeprowadza się na reprezentatywnej i przeszkolonej grupie osób w określonych, kontrolowanych i powtarzalnych warunkach. Osoba oceniająca jakość mowy przydziela jej ocenę według przyjętej skali punktowej, np.: doskonała (5), dobra (4), zadowalająca (3), słaba (2), zła (1), która odpowiada pięciostopniowej skali MOS.

Do najczęściej stosowanych subiektywnych metod badania należą:

- metoda bezwzględnej oceny jakości mowy ACR (*ang. Absolute Category Rating*),
- metoda oceny stopnia degradacji jakości mowy DCR (*ang. Degradation Category Rating*),
- metoda porównawcza oceny jakości mowy CCR (*ang. Comparision Category Rating*),
- metody badania wyrazistości (logatomowej, DRT/MTR, AI, %Alcons).

Każda metoda posiada miarę odniesienia pozwalającą na porównanie i uśrednienie wyników MOS uzyskanych różnymi sposobami.

Pomiary jakości mowy, szczególnie w systemach telekomunikacyjnych, zostały objęte normami, które w większości dotyczą metod subiektywnych. Warunki przeprowadzania badań wyrazistości mowy dla potrzeb telefonii określa Polska Norma PN-90/T-05100 [41]. Badania odsłuchowe w skali MOS opisano w zaleceniu ITU-T P.800 [42]. Normę ta została rozszerzona na kodeki cyfrowe – zalecenie ITU-T P.830 [43]. Metodę porównawczą określania jakości mowy definiuje zalecenie ITU-T P.810: Modulated Noise Reference Unit [44]. Do subiektywnego pomiaru jakości mowy odnoszą się też normy ETSI ETR 250 [45] i ETSI EG 201 377-1 [46].

Metoda bezwzględnej oceny jakości mowy (ACR)

Umożliwia stosunkowo szybki i tani jak na warunki metod subiektywnych bezpośredni pomiar jakości w pięciostopniowej skali. Metoda ACR, jest opisana w zaleceniach ITU-T P.800 [42]. W tej metodzie wykorzystywane są listy testowe złożone z prostych, krótkich, nie związanych z sobą semantycznie zdań. Słuchacze oceniają jakości odsłuchu, wysiłek słuchowego oraz preferowaną głośność. Metoda ACR polega na wyznaczeniu bezwzględnej jakości przekazu głosowego bez zastosowania sygnału odniesienia. Aby wyniki można było uznać za wiarygodne, niezbędna jest wykonanie pomiaru dla licznej ekipy słuchaczy (większej od 12). Wadą metody jest mało precyzyjne określenie tego co tak naprawdę mają oceniać słuchający oraz brak naturalnej zdolności słuchaczy do powtarzalnej oceny takiej cechy, jak „jakość mowy". Dodatkową wadą metody ACR z uwagi na niewielką skalę, jest jej nieczułość na niewielkie zmiany jakości.

Metoda określająca stopień degradacji jakości mowy (DCR)

Pomiar polega na porównaniu wzorcowego sygnału mowy o wysokiej jakości z sygnałem przesłanym przez badany kanał telekomunikacyjny. Metoda DCR opisana została zaleceniach ITU-T P.800 [42]. Dzięki porównaniu z sygnałem wzorcowym metoda ta umożliwia badanie nawet niewielkich zmian jakości mowy. Zadaniem słuchaczy jest określenie stopnia zniekształcenia sygnału mowy w stosunku do sygnału odniesienia w 5-cio stopniowej skali, zmiana: niezauważalna (5), niesłyszalna, ale odczuwalna (4), słabo odczuwalna (3), odczuwalna (2), wyraźnie odczuwalna (1). W efekcie wyznaczany jest tzw. współczynnik degradacyjnej, uśrednionej opinii słuchaczy DMOS (*ang. Degradation Mean Opinion Score*).

Metoda porównawcza oceny jakości mowy (CCR)

Testy przeprowadzane są w takich samych warunkach jak w metodach ACR i DCR jednak słuchaczom nie jest znana kolejność odtwarzania sygnału wzorcowego i testowanego. Dzięki takiej procedurze możliwe jest uzyskanie wyniku testów również w sytuacji gdy sygnał poddawany testowi będzie miał odczuwalną przez słuchacza jakość lepszą niż sygnał wzorcowy. W badaniu tym wyznaczany jest wskaźnik porównawczej, uśrednionej opinii słuchaczy CMOS (*ang. Comparision Mean Opinion Score*)

Badanie wyrazistości logatomowej

Metoda badania wyrazistości logatomowej polega na obliczeniu procentowego stosunku prawidłowo odebranych logatomów do całkowitej liczby nadanych logatomów przez grupę słuchaczy [41]. Logatomy to krótkie jedno-, dwu- lub trzysylabowe wyrazy, które w danym języku nie mają jakiegokolwiek znaczenia. Poprawne rozpoznanie logatomu nie jest zatem wynikiem skojarzenia ze znanym wyrazem ani nie wynika z analizy kontekstowej. Aby dobrze rozpoznać logatom należy poprawnie usłyszeć wszystkie składające się na niego fonemy. Badania mogą być przeprowadzane zarówno w warunkach laboratoryjnych jak i w warunkach naturalnych. Średnią wyrazistość logatomową można obliczyć korzystając ze wzoru:

$$W = \frac{1}{N \cdot L} \sum_{n=1}^{N} \sum_{l=1}^{L} W_{n,l} \, [\%] \tag{5.15}$$

gdzie: $W_{n,l}$ oznacza liczbę poprawnie rozpoznanych logatomów przez n-tego słuchacza z l-tej listy, N – liczba słuchaczy, L - całkowita liczba odczytanych logatomów (zwykle ok. 100 logatomów w jednej liście; w badaniu odczytywane są minimum dwie listy).

Rozrzut wyników pomiarów wyrazistości logatomowej dla całego zbioru słuchaczy można wyznaczyć jako odchylenie średniokwadratowe zgodnie ze wzorem:

$$s = \left[\frac{1}{N \cdot L - 1} \sum_{n=1}^{N} \sum_{l=1}^{L} \left(W_{n,l} - W \right)^2 \right]^{1/2} \tag{5.16}$$

Jeżeli różnica pomiędzy wynikiem dla konkretnego pomiaru a wartością średnią wyrazistości jest większa niż trzykrotność rozrzutu, tj. $\left| W_{n,l} - W \right| > 3s$ - wówczas pomiar ten należy wykluczyć [47].

Metoda ta jest wyjątkowo czasochłonna, wymaga również od słuchaczy przejścia odpowiedniego treningu. Ze względu na fakt, iż duża zrozumiałość mowy nie musi być równoznaczna z jej wysoką jakością, testy wyrazistości nie są wystarczające do całkowitej oceny jakości mowy.

Uproszczone testy DRT/MRT (*ang. Diagnostic/Modified Rhyme Test*) różnią się od metod oceny wyrazistości liczebnością zbioru testowego, brakiem zrównoważenia fonematycznego i strukturalnego materiału testowego oraz możliwością wielokrotnego powtarzania tej samej testowej jednostki dźwiękowej (co jest wykluczone w metodach badających wyrazistość). Metody DRT/MRT można jednak stosunkowo łatwo zautomatyzować [48]. Zbiór jednostek testowych jest najczęściej ograniczony i nieliczny (np. 6 jednostek), natomiast dla danego pomiaru jednostki są prezentowane słuchaczom wielokrotnie w losowej kolejności.

Metoda wskaźnika wyrazistości AI

Wskaźnik wyrazistości AI (*ang. Articulation Index*) jest miarą wyrazistości dla określonego pasma częstotliwości w widmie sygnału. Opiera on się na założeniu, że każdy obszar częstotliwości dźwięku ma procentowo inny udział w procesie rozumienia mowy. Wskaźnik AI wyznaczany jest jako wartość średnia wyrazistości określonych dla poszczególnych pasm częstotliwości zgodnie ze wzorem:

$$AI = \frac{1}{N}\sum_{i=1}^{N} W_i \tag{5.17}$$

gdzie: W_i jest wskaźnikiem wyrazistości w i-tym pasmie elementarnym, N – liczba analizowanych pasm częstotliwościowych.

Jeżeli znana jest zależność wskaźnika wyrazistości W_i od poziomów słyszalności formantów we wszystkich elementarnych pasmach, to można obliczyć całkowity wskaźnik wyrazistości dla danego układu transmisji mowy. Sposób przeprowadzania pomiarów AI opisany został w standardzie ANSI S3.5-1969 [38]. Uaktualnieniem metody AI jest wskaźnik SII (*ang. Speech Inteligibility Index*) opisany w standardzie ANSI (S3.5-1997) [49]. Metoda SII uwzględnia niektóre czynnik (np. pogłos), które nie były uwzględniane w AI.

Metoda %Alcons

Wskaźnik utraty wyrazistości spółgłoskowej $\%Al_{cons}$ (*ang. Articulation Loss of Consonants*) jest miarą dobrze opisaną w literaturze [50] ale nie posiadającą własnego standardu. Polega ona na wyznaczeniu procentowej wartości stosunku liczby błędnie rozpoznanych przez słuchacza spółgłosek do liczby wszystkich spółgłosek przesłanych przez danych układ. Za transmisję o doskonałej jakości uważa taką, w której $Alcons \leq 2\%$. Natomiast maksymalna dopuszczalna wartość nie może przekraczać 15%.

Metoda ta choć jest prosta i dość powszechnie stosowana (szczególnie przez konsultantów akustycznych) prowadzi do błędnych wyników w warunkach pogłosowych lub przy obcinaniu szczytów sygnałów [11].

Pomiar zrozumiałości mowy na tle szumu

Do subiektywnych badań zrozumiałości na tle szumów wykorzystywanych jest wiele różnych testów różniących się między sobą zawartością materiału testowego (zdania lub słowa) oraz parametrami poziomu natężenia i rodzaju szumu zakłócającego. W badaniach wyznaczany jest próg percepcji mowy SRT (*ang. Speech Reception Threshold*), który określa minimalny stosunek sygnału do szumu umożliwiający zrozumiałość na poziomie 50%. Do porównania różnych testów stosuje się funkcje zrozumiałości opisane wzorem:

$$p(L, STR, s) = \frac{1}{1 + e^{4s(SRT - L)}} \tag{5.18}$$

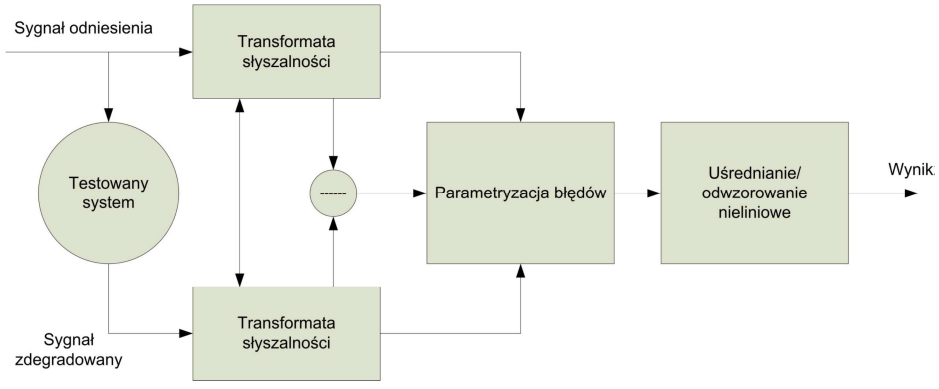
gdzie: p – prawdopodobieństwo prawidłowego rozpoznawane przy stosunku sygnału do szumu równym L, s - nachylenie wykresu zrozumiałości.

W testach wykorzystywane mogą być testy zdaniowe jak np. zdania typu Plomp'a [51], polski test zdaniowy [52], polski test typu Matrix [53]) jak również listy logatomowe [41] lub cyfrowe (polski test trypletów cyfrowych [54]).

Testy zadaniowe z uwagi na przenoszenie informacji kontekstowych pomagających słuchaczowi ich rozpoznanie zazwyczaj mają bardziej pochylone zbocza funkcji psychometrycznych czyli są mniej wrażliwe na zmiany poziomu szumu. Testy wyrazowe i logatomowe posiadają zazwyczaj niższą wartość SRT ale jednocześnie odseparowują badanie od efektu kontekstowego i pozwalają rzetelniej ocenić degradację sygnału związaną wyłącznie z parametrami transmisyjnymi kanału.

5.4. Metody obiektywnej oceny jakości przekazu głosowego

Podstawową zaletą obiektywnych metod badania jakości przekazu mowy jest to, że są szybsze, prostsze i tańsze w stosunku do metod subiektywnych. Opierają się ona na wyznaczeniu odległości, czyli pewnej miary, pomiędzy wartościami wybranych parametrów (jednego lub kilku) sygnału wzorcowego oraz sygnału badanego (np. zniekształconego w wyniku błędnej transmisji). Koncepcję metody porównawczej oceny jakości mowy (rys. 5.3) zaproponował Karjalainen w 1985 r. [55].

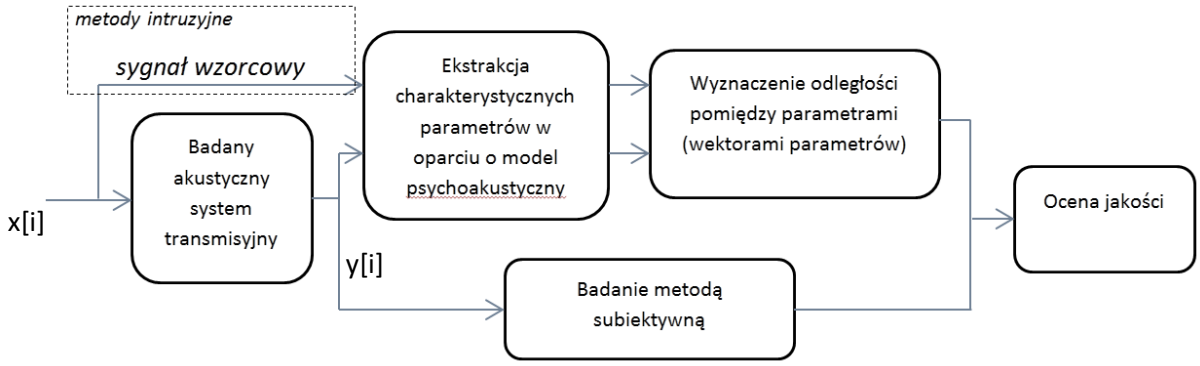


Rysunek 5.3 Ogólna koncepcja metody porównawczej szacowania jakości mowy [55].

Wzorcowy sygnał odniesienia jest wykorzystywany w metodach intruzyjnych natomiast w testach nieintruzyjnych analizowane są jedynie parametry kanału transmisyjnego, w tym zakłócenia, szumy, nieciągłości.

Po wyznaczeniu miary odległości sygnału oryginalnego i zdegradowanego, na podstawie krzywej odniesienia następuje przeskalowanie jej wartości do subiektywnej miary odniesienia

(np. MOS). Na rysunku 5.4 przedstawiono uproszczony schemat blokowy obiektywnej oceny jakości opartej na analizie parametrów sygnału mowy.



Rysunek 5.4 Uproszczony schemat obiektywnej oceny jakości mowy; *x[i]* – wzorcowy sygnał odniesienia; *y[i]* – sygnał zniekształcony.

5.4.1. Jakości sygnału mowy transmitowanej w systemach w telekomunikacyjnych

Do badania jakości sygnału mowy transmitowanej w systemach w telekomunikacyjnych opracowano wiele odpowiednich testów obiektywnych. Do najczęściej stosowanych należą:

- PSQM (*ang. Psycho-Acoustic Speech Quality Measure*) - opisany w zaleceniu ITU-T P.861 [56],
- PESQ (*ang. Perceptual Evaluation of Speech Quality*) - poprawiona metoda PSQM - opisany została w zaleceniu ITU-T P.862 [57],
- PEAQ (*ang. Perceptual Evaluation of Audio Quality)* – norma BS.1387 [58] – opracowana dla sygnałów akustycznych w pełnym paśmie częstotliwości odbieranym przez człowieka.

Miary PSQM, PESQ, PEAQ należą do grupy metod intruzyjnych, uwzględniają również niektóre aspekty psychoakustyczne, np. maskowanie zakłóceń.

Metoda PSQM

Metoda PSQM przeznaczona jest do badania systemów transmisji telefonicznej w ograniczonym paśmie od 300 do 3400 Hz zgodnie z zaleceniem ITU-T P.861 [56]. Wyniki testów przeprowadzanych tą metodą posiadają najwyższą korelację z wynikami testów subiektywnych, sięgającą nawet 98%. Mimo tego, że metoda ta jest stosunkowo prosta nie uzyskała akceptacji do badań jakości mowy transmitowanej w sieci komputerowych VoIP (*ang. Voice over Internet Protocol*) gdyż nie uwzględnia ona wpływu utraty oraz zmiennego opóźnienia pakietów danych.

Testy jakości mowy w oparciu metodę PSQM polegają na podobnie jak wszystkie testy obiektywne na porównywaniu wzorcowego sygnału wejściowego *x[i]* z wyjściowym sygnałem *y[i]* zniekształconym w wyniku transmisji przez badany układ.

Przed przystąpieniem do porównywania wewnętrznych reprezentacji sygnałów poddawane są one przekształceniom odpowiadającym psychofizycznemu odbiorowi mowy przez ludzkie ucho. Percepcja mowy jest zróżnicowana w zależności od częstotliwości i względnego poziomu głośności. Transformacja sygnałów z wykorzystaniem modelu percepcyjnego z postaci fizycznej na postać psychofizyczną odbywa się w trzech etapach:

- przekształcenie czasowo-częstotliwościowe – poprzez wyznaczenie krótkookresowej, *N*-punktowej dyskretnej transformaty Fouriera DFT dla fragmentu N-próbek sygnału ramkowanego oknem Hanninga (najczęściej *N*=256 dla *fs*=8kHz),
- przeskalowanie częstotliwości - za pomocą filtrów barkowych, w których szerokości pasma oraz częstotliwość środkowa filtru zwiększają się wraz z częstotliwością. Na sygnał nakładany jest dodatkowo maskujący szum Hotha,
- przeskalowanie poziomu natężenia - w celu uwzględnienia subiektywnego odczucia głośności wykonywane jest przeskalowanie poziomu natężenia ze skali decybelowej na fonową (gdyż subiektywne odczucie głośności nie zależy liniowo od poziomu natężenia sygnału wyrażonego w decybelach) a następnie wyraża się tą głośność w skali sonowej (gdyż subiektywne wrażenie zwiększenia głośności zmienia się nieliniowo wraz ze zmianą głośności).

W wyniku porównania wewnętrznej reprezentacji tak przekształconych sygnałów (wzorcowego i zdegradowanego) wyznaczane są odległości pomiędzy ich parametrami oraz obliczany jest wskaźnik PSQM, który jest bezpośrednio powiązany z jakością badanego sygnału mowy. Wskaźnik ten przyjmujący wartości w skali od 0 (jakość doskonała) do 6.5 (bardzo niska) może zostać następnie przekształcony do subiektywnej miary MOS.

Metoda PESQ

Metoda PESQ jest rozwinięciem testu PSQM ale wykorzystuje odmienny model percepcyjny oraz inny sposób przetwarzania sygnałów. Metoda ta opisana została w zalecenie ITUT P.862 [57]. W stosunku do metody PSQM dodano w niej algorytm wyrównywania czasowego sygnałów oraz wyrównywania mocy pomiędzy dwoma Sygnałami, dzięki czemu zaadaptowano ją do pomiarów jakości przekazu mowy w sieciach VoIP. Pozostała cześć badania przebiega podobnie jak w metodzie PSQM. Dla sygnałów wzorcowego i testowanego wyznaczane są ich reprezentacje wewnętrzne, a następnie obliczane są między nim różnice odpowiadające bezpośrednio jakości mowy. Wskaźnik wynikowy PESQ zawiera jest w skali oceny od -0,5 do 4,5, który jest następnie kalibrowany według skorelowanej skali MOS od 1 do 5 zgodnie z ITU-T P.800 [42].

Podstawową zaletą tej metody jest uwzględnienie większości czynników degradujących jakość mowy. Metoda testów PESQ wykorzystywana jest głównie przy badaniach i projektowaniu kodeków cyfrowych dla sieci IP gdyż wykazuje się największą spośród stosowanych w takich sieciach metod korelacją z wynikami testów subiektywnych, rzędu 90-95%.

Metoda PEAQ

Stosowana dla szerokopasmowego sygnału audio zgodnie z zaleceniem BS.1387 [58]. W metodzie PEAQ porównuje się oryginalny sygnał referencyjny do sygnału po zakodowaniu oraz do sygnału odkodowanego. Procedura przekształcenia sygnałów w modelu percepcyjny jest analogiczna do tej opisanej dla metody PSQM. Po przeskalowaniu sygnałów i wyznaczeni FFT, uwzględnia wpływ ucha zewnętrznego i środkowego oraz zjawiska maskowania następuje wyodrębnienie cech wewnętrznych sygnałów i obliczenie wskaźnika jakości na podstawi różnicy pomiędzy wartościami cech sygnału.

5.4.2. Zrozumiałości mowy transmitowanej w pomieszczeniach pogłosowych

Sygnał mowy transmitowany w układzie złożonym z systemu nagłośnieniowego rozmieszczonego w pomieszczeniu pogłosowym poddawany jest dużym zniekształceniom. Dążenie do zapewnienia dostatecznego poziomu natężenia dźwięku w miejscu odsłuchu powodować może znaczne pogorszenie zrozumiałości przekazu z uwagi na silne wielokrotne odbicia (rewerberacje) generujące silny pogłos.

Do pomiaru zrozumiałości mowy w takich warunkach wykorzystywane są wskaźniki parametryzujące zjawiska pogłosowe takie jak: krzywa pogłosowa, czas wczesnego zaniku, współczynnik dźwięku bezpośredniego do pogłosowego, wyrazistość, przejrzystość, odstęp echa, wskaźnik pogłosu i inne oraz parametry przestrzenne takie jak: skuteczność odbić bocznych czy reakcja pomieszczenia.

Dugą grupą metod oceny zrozumiałości w warunkach pogłosowych są metody oparte na funkcji przeniesienia modulacji MTF (*ang. Modulation Transfer Function*) takie jak STI, RASTI, STIPA.

Metody te opierają się na założeniu, iż w wyniku zniekształceń pogłosowych następuje zmniejszenie głębokości modulacji transmitowanego przez układ szumu testującego. Badania zmian modulacji dokonuje się w poszczególnych pasmach oktawowych dla częstotliwości modulujących wynikających z podzielenia przedziału 0,63 - 12,5 Hz na pasma tercjowe, odpowiadające obwiedni sygnału mowy naturalnej. Modulacji poddawany jest szum posiadający średnią widmową gęstość mocy mowy naturalnej, przy czym współczynnik modulacji sygnału nadawanego wynosi m=100 %.

Metoda oceny zrozumiałości STI (*ang. Speech Transmission Index*), wprowadzona w 1971 r. [12] opiera się na pomiarach wykonywanych w 7 pasmach oktawowych, których częstotliwości środkowe z zakresu 125 - 8000 Hz są modulowane czternastoma różnymi częstotliwościami modulującymi z przedziału 0,63 - 12,5 Hz. Metoda ta jest czasochłonna i kosztowna obliczeniowo, wymaga bowiem wykonania 98 pomiarów w każdym punkcie badanego obszaru. Metoda obliczania wskaźnika opisana została w IEC 60268-16 [11].

Wyliczone wartości wskaźnika zrozumiałości prezentowane są w wartościach skali CIS zgodnie z zaleceniem IEC 60849 [40] dla systemów ostrzegawczych.

Wskaźnik transmisji mowy STI obliczany jest zgodnie z zależnością [59] [60]:

$$STI = \frac{1}{w}\sum_{j=1}^{L} w_j MTF_j \tag{5.19}$$

gdzie: MTF_j - wskaźnik funkcji przeniesienia modulacji dla *j*-tego podpasma oktawowego, w_j - współczynnik wagowy dla *j*-tego podpasma oktawowego.

Możliwe jest również wyznaczenie MTF z odpowiedzi impulsowej układu akustycznego [61], w oparciu o następującą zależność:

$$MTF(F_{\mathrm{mod}}) = \left| \frac{\sum_{i=kd}^{kg} H(i) \cdot H(L_F - i)}{\sum_{i=kd}^{kg} \left| H(i)^2 \right|} \right| \tag{5.20}$$

gdzie: L_F – numer prążka odpowiadający częstotliwości modulującej F_{mod}.

W celu zredukowania czasochłonności procesu pomiarowego STI opracowano jego uproszoną wersję RASTI (*ang. RApid Speech Transmission Index*), która wykorzystuje tylko 9 pomiarów w dwóch pasmach oktawowych o częstotliwościach środkowych 500 Hz i 2000 Hz z częstotliwościami modulującymi z zakresu 1-8 Hz (dla pasma $f_{śr}$=500 Hz) oraz 0,7 - 11,2 Hz (dla pasma $f_{śr}$=2000 Hz).

Ponieważ metoda RASTI daje tylko przybliżony wynik a pełna metoda STI jest bardzo czasochłonna opracowano również metodę pośrednią STIPA (ang. *Speech Transmission Index for Public Address Systems*) wykorzystującą 7 pasm oktawowych z 12 częstotliwościami modulującymi.

Pomiary STI dają dobre rezultaty jedynie w odniesieniu do układów liniowych w których nie są stosowane takie urządzenia jak wokodery, procesory efektów czy układy przesuwające lub powielające częstotliwość.

5.5. Miary zniekształceń sygnału mowy oparte na liniowej predykcji

Dla każdego fonemu, stan traktu głosowy mówcy daje się opisać zbiorem parametrów, które znajdują odzwierciedlenie w widmie transmitancji filtru traktu głosowego w postaci występujących w nim charakterystycznych maksimów tzw. formantów. Położenie oraz poziom formantów występujących w fonemach wpływa na możliwości ich rozróżnienia przez słuchacza. Skutkiem interferencji szerokopasmowej są zmian parametrów sygnału mowy, które wygodnie jest interpretować właśnie jako zmiany parametrów filtru traktu głosowego. Podejście takie umożliwia wyznaczenie obiektywnych wskaźników odległości pomiędzy wektorami współczynników filtru dla sygnału oryginalnego i zmienionego, które to wskaźniki można traktować jako mierniki stopnia zniekształcenia sygnału względem jego pierwotnej postaci.

Przyjęta przez autora metoda pomiaru stopnia zniekształceń sygnałów mowy opiera się na obliczeniu dla sygnałów: oryginalnego i zniekształconego wartości odległości pomiędzy zbiorami parametrów predykcji liniowej (LPC) oraz w celach porównawczych współczynników cepstralnych i mel-cepstralnyh (MFCC. Do obliczeń wykorzystano znane metryki badania odległości.

Jedną z metryk jest miara LLR (*ang. Log-Likelihood Ratio*). Pomiar odległości LR *(ang. Likelihood Ratio*), który został wprowadzony przez F. Itakura [62] [63] [64], wykorzystuje do ilościowego opisu zniekształceń w widmie sygnału mowy współczynniki LPC parametryzujące obwiednię widma zarówno sygnału wejściowego jak i wyjściowego (zniekształconego). Zniekształcenia te standardowo związane są z procesem kodowanie mowy w systemach telekomunikacyjnych ale dzięki ich uniwersalnym właściwościom możliwe jest ich zastosowanie do oceny deformacji sygnału innych typów, w tym również zniekształceń wprowadzanych przez interferencję szerokopasmową, co zostało przez autora rozprawy pokazane w [18] [17] [16].

Odległość LR jest zdefiniowana jako stosunek energii sygnałów resztkowych przed i po dokonaniu operacji kodowania LPC. Wynikowe współczynniki predykcji $a_{IN}=\{a_0, a_1, \ldots, a_p\}$ są wyznaczane tak aby minimalizować energię sygnału resztkowego (4.4). Każda zmiana wektora współczynników powoduje więc zawsze zwiększenie poziomu energii sygnału resztkowego. Estymacja sygnału w oparciu o wartości elementów wektora a_{IN}, które w wyniku zniekształceń sygnału przyjmują nowe wartości $a_{OUT}=\{a_0, a_1, \ldots, a_p\}$ prowadzi do syntezy sygnału będącego zdeformowaną wersją sygnału oryginalnego. Zdekodowanie sygnału w oparciu o błędne współczynniki predykcji spowoduje więc zawsze wzrost błędu predykcji, a co za tym idzie wzrost energii sygnału resztkowego. Wynika więc z tego, że wartości współczynnika LR >=1.

W wyniku zlogarytmowania współczynnika LR otrzymuje się miarę odległości LLR, którą można zdefiniować jako:

$$d_{LLR}(\mathbf{a}_d, \mathbf{a}_c) = \log\left(\frac{\mathbf{a}_d^T \mathbf{R}_c \mathbf{a}_d}{\mathbf{a}_c^T \mathbf{R}_c \mathbf{a}_c}\right) \tag{5.21}$$

gdzie $\mathrm{a_c}$ jest wektorem współczynników LPC sygnału niezakłóconego, $\mathrm{a_d}$ - wektorem współczynników LPC sygnału zakłóconego, $\mathrm{R_c}$ natomiast to macierz autokorelacji sygnału niezakłóconego.

Miara odległości zdefiniowana w (5.21) nie jest symetryczna, nie spełnia więc warunków metryki (5.1). Problem ten może zostać rozwiązany poprzez zastosowanie symetrycznej transformacji w postaci:

$$d_{LLRS}(\mathbf{a}_d, \mathbf{a}_c) = \frac{d_{LLR}(\mathbf{a}_d, \mathbf{a}_c) + 1/d_{LLR}(\mathbf{a}_d, \mathbf{a}_c)}{2} - 1 \tag{5.22}$$

Kolejną miarą opartą na metodzie liniowej predykcji jest miara zniekształceń Itakura-Saito (IS). Zgodnie z (5.10), przyjmując jako wektor cech sygnału energię sygnału resztkowego związaną z błędem predykcji otrzymujemy zależność:

$$d_{IS}(\mathbf{a}_d, \mathbf{a}_c) = \left[\frac{\sigma_c^2}{\sigma_d^2}\right]\left[\frac{\mathbf{a}_d \mathbf{R}_c \mathbf{a}_d^T}{\mathbf{a}_c \mathbf{R}_c \mathbf{a}_c^T}\right] + \log\left(\frac{\sigma_c^2}{\sigma_d^2}\right) - 1 \tag{5.23}$$

gdzie σ_c^2 i σ_d^2 to wariancje błędu predykcji odpowiadająca wzmocnienia toru dla sygnału odpowiednio: wzorcowego i zakłóconego.

Jak już wspomniano wcześniej odległość IS nie jest symetryczna, a jej symetryczną realizacją jest metryka „Cos-h" zgodnie z (5.11). Bierze ona pod uwagę ogólny poziom obwiedni widma sygnału, która nie odzwierciedla w pełni psychoakustycznych właściwości słuchu.

5.6. Wybór metodyki i zastosowanych miar

Aby dokonać prawidłowej oceny jakości przekazu sygnału należy zastosować metodę pomiaru adekwatną do rodzaju badanych sygnałów oraz dostosowaną do charakteru zniekształceń. Metoda badania implikuje wybór niezbędnych parametrów opisujących parametry sygnału i kanału komunikacyjnego, a także zastosowane wskaźniki oraz miary różnic pomiędzy wartościami wskaźników. Inny charakter mają zakłócenia w transmisji sygnału mowy w sieciach telekomunikacyjnych a inny zniekształcenia w pomieszczeniach pogłosowych.

W wyniku przeprowadzonego przeglądu podstawowych metod badania jakości przekazu mowy nie dokonano wyboru żadnej ze stosowanych metod jako dobrze odzwierciedlającej charakter zniekształceń powstających w polu zespołów źródeł szerokopasmowych. Stosowane powszechnie obiektywne metody pomiarów i oceny zrozumiałości mowy w pomieszczeniach np. STI, RASTI, jak to pokazano w niniejszej pracy, są niewrażliwe na wpływ zjawiska superpozycji sygnałów dochodzących do słuchacza z wielu źródeł.

W związku z powyższym zdecydowano się na wybranie metody badania odnoszącej się bezpośrednio do zauważalnych zmian obiektywnych parametrów związanych z mechanizmem generowania głosu powstających w skutek interferencji szerokopasmowej. Przyjęta metoda pomiaru stopnia zniekształceń sygnałów mowy opiera się na obliczeniu dla sygnałów: oryginalnego i zniekształconego wartości odległości pomiędzy zbiorami parametrów predykcji liniowej (LPC) oraz w celach porównawczych współczynników cepstralnych i mel-cepstralnyh (MFCC).

Do obliczeń wykorzystano matematyczne metryki stosowane do wyznaczania odległości pomiędzy wektorami. Z przeprowadzonych testów wynika, iż najbardziej efektywne wyniki daje wskaźnik IS, którego wartość jest definiowana w oparciu o stosunek energii sygnałów resztkowych przed i po dokonaniu operacji kodowania LPC, w którym to wynikowe współczynniki predykcji są wyznaczane tak aby minimalizować energię sygnału resztkowego. Każde zniekształcenie sygnału, a co za tym idzie zmiana wektora współczynników, powoduje zawsze zwiększenie poziomu energii sygnału resztkowego. Estymacja sygnału w oparciu o nowe wartości elementów wektora, prowadzi do syntezy sygnału będącego zdeformowaną wersją sygnału oryginalnego co determinuje wzrost błędu predykcji i energii sygnału resztkowego. Implikuje to także wzrost wartości wskaźnika odległości dzięki czemu jest on dobrym miernikiem istotności zmian w analizowanych sygnałach mowy.

6. WYNIKI BADAŃ SYMULACYJNO-POMIAROWYCH PARAMETRÓW SYGNAŁU MOWY

Poniżej przedstawiono wybrane wyniki badań symulacyjnych i pomiarowych dla trzech typowych układów wielośródłowych odpowiadających rzeczywistym realizacjom systemów nagłaśniających, tj. układ szyku źródeł rozłożonych w jednej linii, układ ciągu komunikacyjnego oraz układ matrycowy (sala audytoryjna). W obszarach odsłuchowych każdego z tych układów wyznaczone zostały następujące wskaźniki odległości: LLR (Log-Likelihood Ratio), IS (Itakura-Saito), CD (Cepstrum Distance), melCD (MFCC Distance). Jako potwierdzenie istotnego wpływu interferencji szerokopasmowej na jakość i zrozumiałość przekazu głosowego przedstawiono wyniki przeprowadzonego testu komputerowego rozpoznawalności polskich trypletów cyfrowych, transmitowanych w systemie wielośródłowym.

6.1. Zobrazowanie rozkładu zmian wskaźników odległości w polu akustycznym układów wielośródłowych

Wartości wskaźników LLR, IS, CD i melCD (zobrazowanych na wykresach zaprezentowanych w dalszej części rozdziału) obliczone zostały pierwotnie w oparciu o symulacje komputerowe idealnych (teoretycznych) odpowiedzi impulsowych, wyznaczonych na podstawie zależności geometrycznych. Te same wskaźniki wyznaczono następnie przy uwzględnieniu rzeczywistych odpowiedzi impulsowych, pomierzonych metodą korelacyjną opisaną w rozdz. 3.1. Charakterystyki tak pomierzonych rzeczywistych źródeł (głośników) przedstawione są na rysunkach 3.6 i 3.7. Obliczenia teoretyczne wykonane zostały przy zastosowaniu aplikacji opisanej w rozdz. 3.3.

Do badań przyjęto upraszczające założenie o dookólnych charakterystykach kierunkowych dla całego badanego widma, zarówno teoretycznych jak i rzeczywistych źródeł. Dla źródeł teoretycznych założono, że posiadają one płaską funkcję przenoszenia dla całego zakresu częstotliwości emitowanego sygnału. Źródła rzeczywiste posiadają natomiast funkcje przenoszenia pokazane na rys. 3.7. We wszystkich obliczeniach przyjęto również założenie oraz braku odbić i pochłaniania dźwięku (pole swobodne, bez przeszkód pomiędzy źródłem a punktem odsłuchu).

Dla wszystkich prezentowanych w dalszej części rozdziału wykresów prezentujących mapy zmienności wskaźników odległości, zastosowano wspólną skalę kolorów. Kolor niebieski odpowiada najniższym wartościom wskaźników i wskazuje obszary, w których brak jest zniekształceń lub są one niewielkie. Kolory od żółtego do czerwonego wskazują na obszary o znacznym stopniu zniekształcenia względem sygnału oryginalnego.

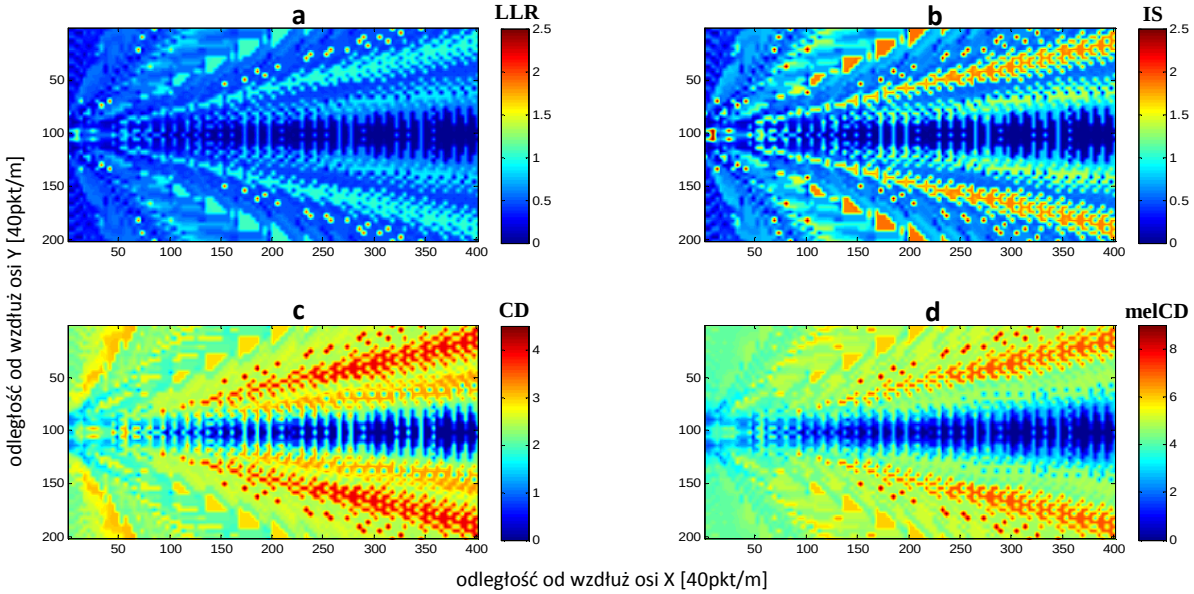
W dalszej części rozdziału zaprezentowano jedynie wybrane przykłady map zmienności, dla wybranych układów i fonemów, które szczególnie dobrze prezentują charakterystykę zmian wartości obliczonych wskaźników. Wyniki analizy tych samych układów, ale dla innych

rozkładów przestrzennych źródeł oraz punktów odsłuchu, zostały przedstawione szerzej w Dodatkach B1-B3.

6.1.1. Układ typu szyk źródeł w jednej linii

Poniżej przedstawiono zmienność wskaźników wyznaczonych dla sygnałów głosowych transmitowanych przez układ źródeł, określany w dalszej części jako typ szyku źródeł w jednej linii (5 źródeł rozłożonych równomiernie co d=0.2m wzdłuż linii). Płaszczyzna odsłuchu (5m x 10m), na której badane są wartości wskaźników, znajduje się w płaszczyźnie osi apertury.

Porównanie map zmienności dwóch wskaźników, IS oraz melCD (rys. 6.1b i 6.1d oraz 6.2 i B1.1), wskazuje na sporą różnicę w dynamice ich zmian. Jednocześnie względne zmienności wartości wskaźników IS oraz melCD w całym badanym obszarze są niemal identyczne. Taka zgodność daje dużą dozę pewności co do prawidłowości ich zastosowania oraz daje możliwość ich zamiennego stosowania. Dwa pozostałe wskaźniki używane przez autora, tj. LLR oraz CD, również wykazują niemal identyczny rozkład zmienności na badanej płaszczyźnie w stosunku do rozkładu zmian wskaźnika IS (rys.6.1a i 6.1c). Z uwagi na zalety wskaźnika IS, który jest wygodniejszy do stosowania od pozostałych gdyż lepiej uwidacznia zróżnicowanie zmian wartości w całym zakresie ich zmienności, został przez autora przyjęty jako podstawowy miernik jakości, odwzorowujący stopnień zniekształcenia sygnału głosowego.

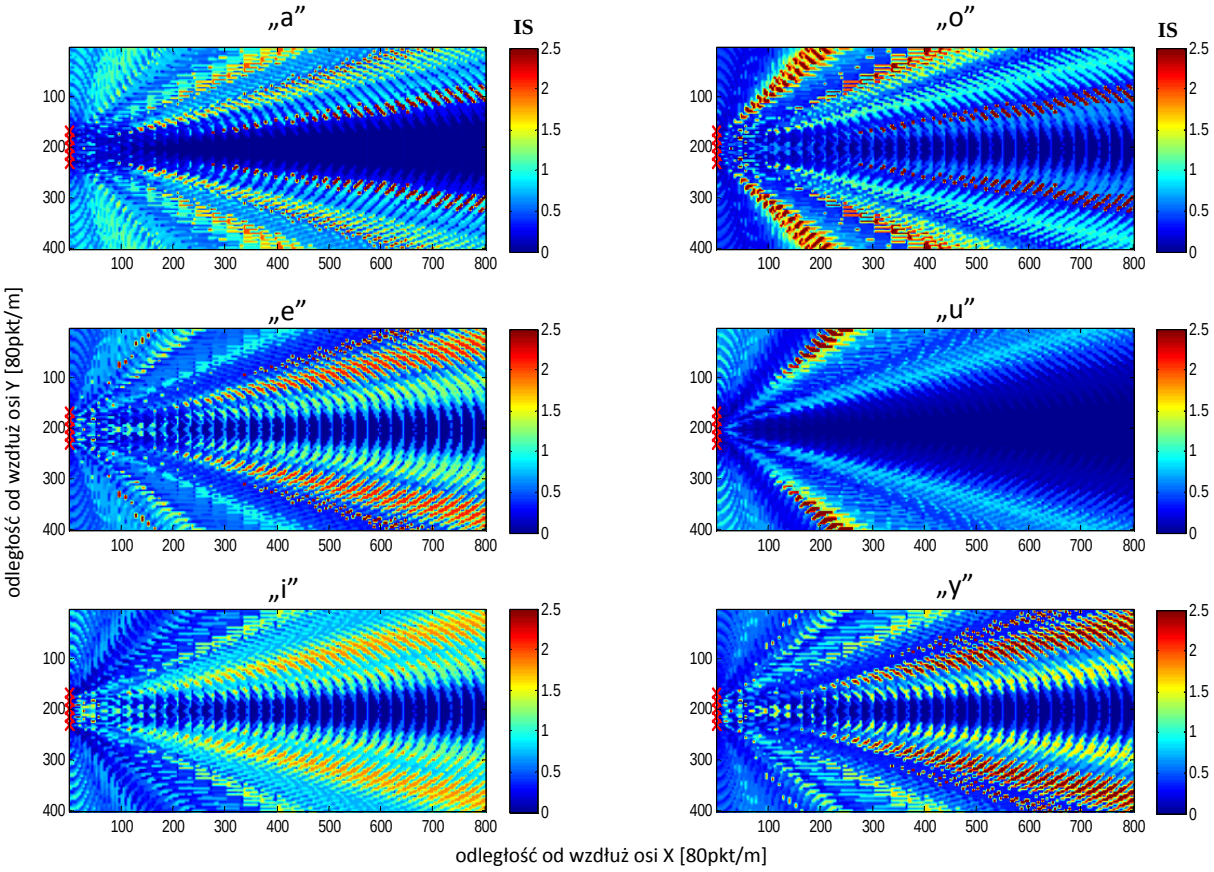


Rysunek 6.1 Porównanie map zmienności wskaźników odległości: a) LLR; b) IS; c) CD; d) melCD w funkcji położenia punktu odsłuchu, dla przykładowego układu typu szyk źródeł w jednej linii, dla głoski „e”.

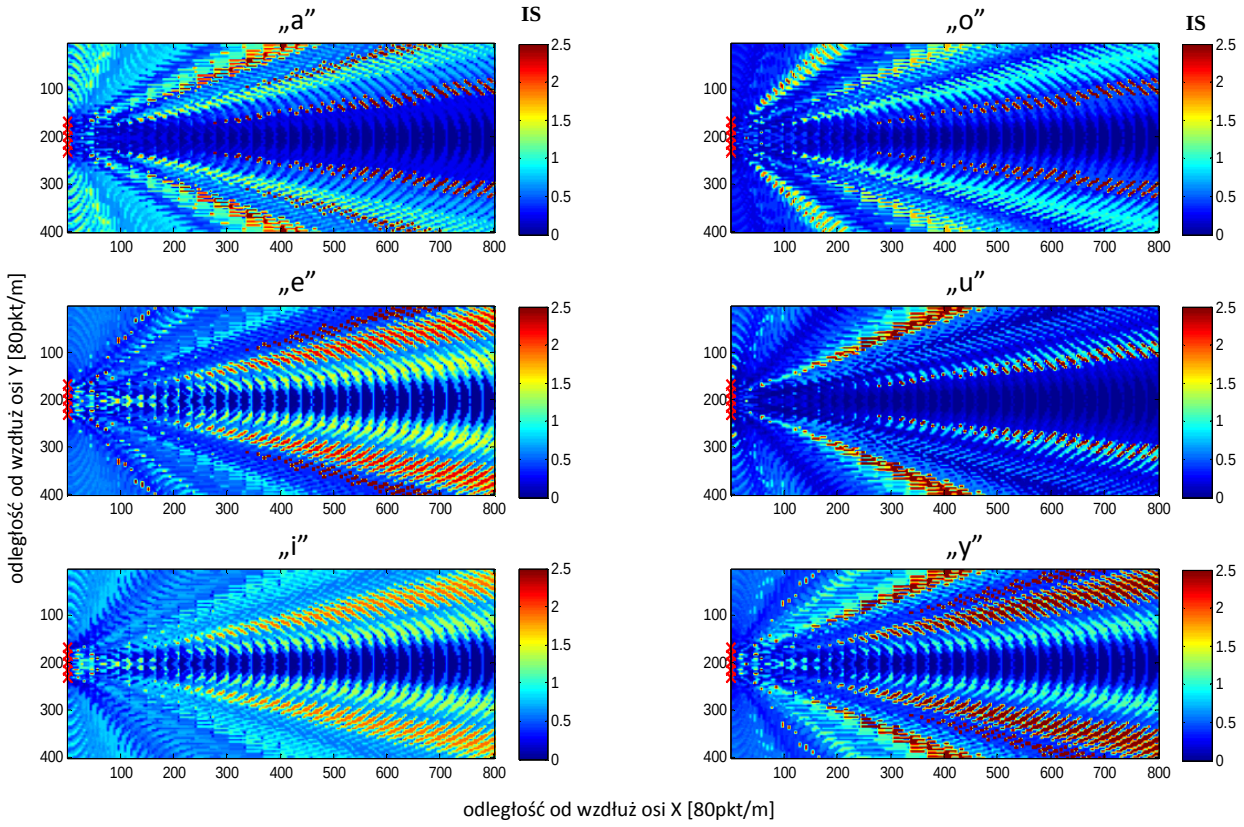
Na rysunku 6.2 przedstawione zostały mapy zmienności wskaźnika IS dla sześciu samogłosek: „a”, „e”, „i”, „o”, „u”, „y”, transmitowanych przez układ źródeł typu szyk w jednej linii. W obszarze odsłuchu dla każdej z głosek zauważalna jest duża zmienność wartości wskaźników.

Obszary, w których występują stosunkowo małe zniekształcenia są rozdzielone obszarami, w których zauważalny jest duży stopień zniekształcenia sygnału mowy.

Obliczenia wykonane dla tego samego układu jak na rys. 6.2, ale wyznaczone z uwzględnieniem odpowiedzi impulsowych rzeczywistych źródeł (głośników), pokazano na rysunku 6.3.

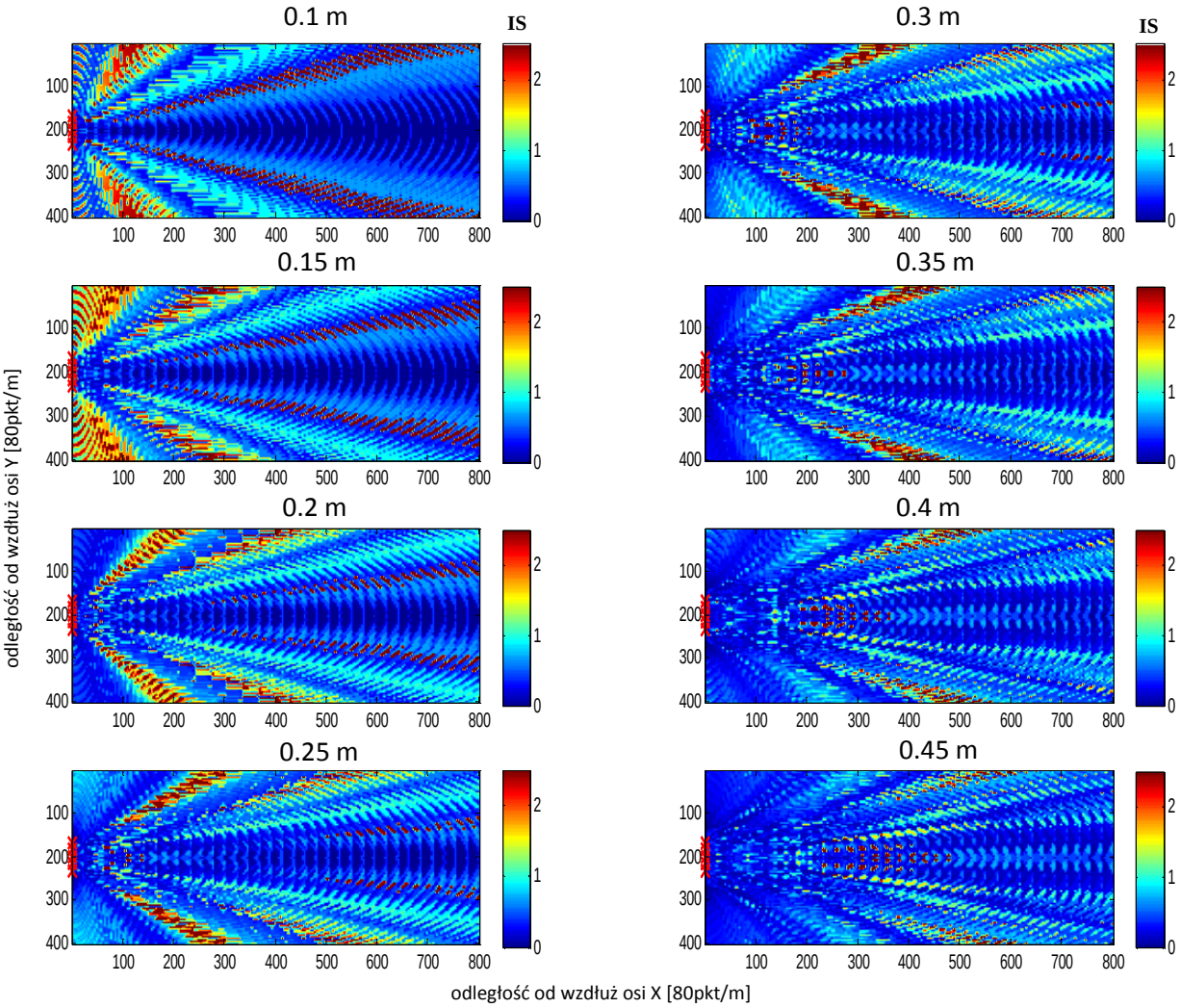


Rysunek 6.2 Mapy zmienności wskaźnika IS dla sześciu samogłosek, dla liniowego układu 5 źródeł typu szyk źródeł w jednej linii, rozstaw źródeł 0.2 m (źródła idealne).



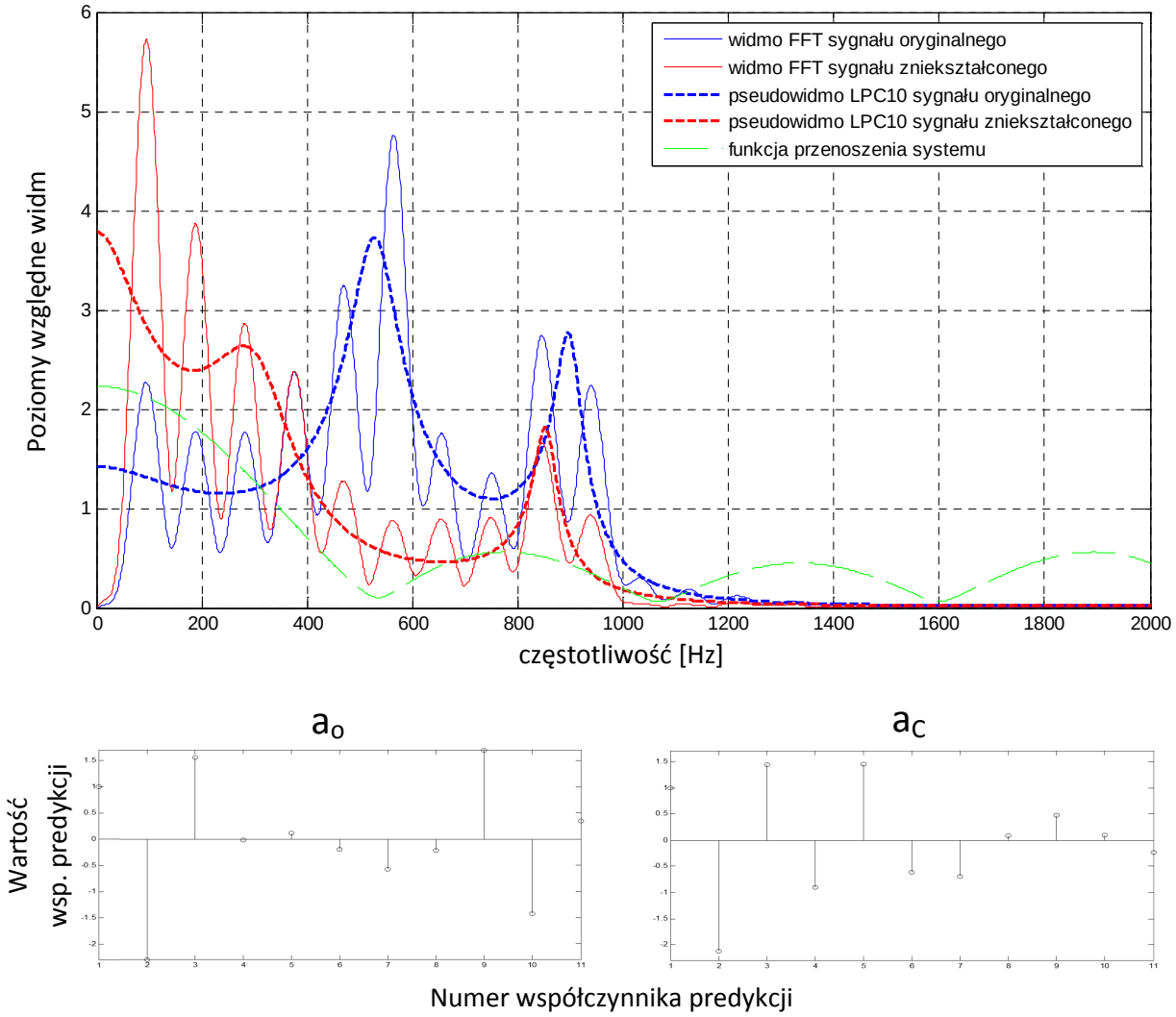
Rysunek 6.3 Mapy zmienności wskaźnika IS dla sześciu samogłosek, dla liniowego układu 5 źródeł typu szyk źródeł w jednej linii, rozstaw źródeł 0.2 m. Badane fonemy, układy źródeł oraz położenia płaszczyzny odsłuchu są identyczne jak na rys. 6.2 ale obliczenia wykonano dla odpowiedzi impulsowych rzeczywistych źródeł (głośników).

Na rysunku 6.4 pokazano w jaki sposób zmienia się wartość wskaźnika IS dla głoski „o” w funkcji zmiany rozstawu źródeł w zakresie od d=0.1m do d=0.45m. Wraz ze wzrostem rozstawu źródeł zaobserwować można wyrównanie wartości wskaźników przy jednoczesnym zmniejszeniu poziomu zniekształceń w całym badanym obszarze. Podobny efekt uzyskano również dla innych głosek (rys. B1.2).



Rysunek 6.4 Mapy zmienności wskaźnika IS dla głoski „o” dla liniowego układu 5 źródeł typu szyk źródeł w jednej linii dla ośmiu rozstawów źródeł w zakresie od 0.2 m do 0.45 m (źródła idealne).

Na rysunku 6.5 zaprezentowano zmiany charakterystyk częstotliwościowych (FFT i LPC10) oraz zmiany wartości współczynników predykcji dla samogłoski „o”, w przykładowo wybranych punktach odsłuchu, o stosunkowo dużych wartościach współczynnika IS (dla układu typu szyk w jednej linii). Dużej wartość współczynnika IS odpowiadają istotne zmiany charakterystyk częstotliwościowych transmitowanego sygnału. W poniższym przykładzie (rys. 6.5) wyraźnie widoczna jest filtracja w okolicach częstotliwości pierwszego formantu. Podobne filtrujące odziaływanie układów wielośródłowych zaobserwowano również w przypadku analizy widm innych głosek (rys. A.4).

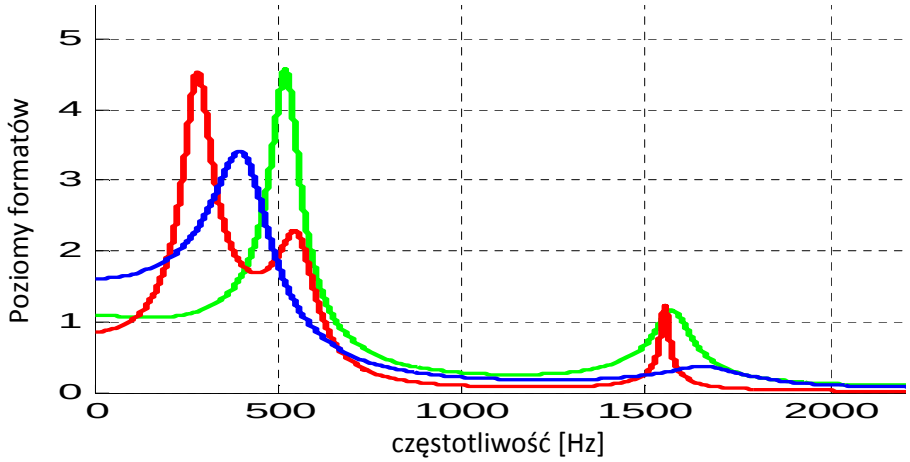


Rysunek 6.5 Zniekształcenia widma FFT oraz pseudowidma LPC10 sygnału (głoska „o"), dla układu typu szyk źródeł w jednej linii (5 źródeł, rozstaw 0.2m - źródła idealne), w punkcie odsłuchu x=1.85m, y=0.95m , kąt = 40 st. Wartość wskaźnika IS = 1.82. Poniżej wykresów zamieszczono wartości współczynników predykcji sygnału oryginalnego a_0 i zniekształconego a_C.

Na uwagę zasługuje fakt silnej zmienności charakterystyk formatowych w miarę zmiany położenia punktu pomiaru. Przesuwając go wzdłuż pewnej linii można wyznaczyć charakterystyki zmian pseudowidma LPC, w postaci zobrazowań dwu- lub trójwymiarowych (przykłady zaprezentowano w dodatku A na rys. A.1-3). Sytuacja taka odpowiada ruchowi słuchacza w obszarze odsłuchu, co jest szczególnie charakterystyczne dla układów akustycznych typu „ciąg komunikacyjny", opisanych w dalszej części rozdziału. W przypadku takim słuchacz może odczuwać znaczne zmiany brzmienia docierających do niego sygnałów głosowych spowodowane silnymi zmianami położenia i poziomów formantów.

Zmiany charakterystyk formatowych mogą następować bardzo szybko, przy nawet nieznaczny przesunięciu punktu odsłuchu. Na rysunku 6.6 przedstawiono pseudowidmo LPC10 dla dwóch

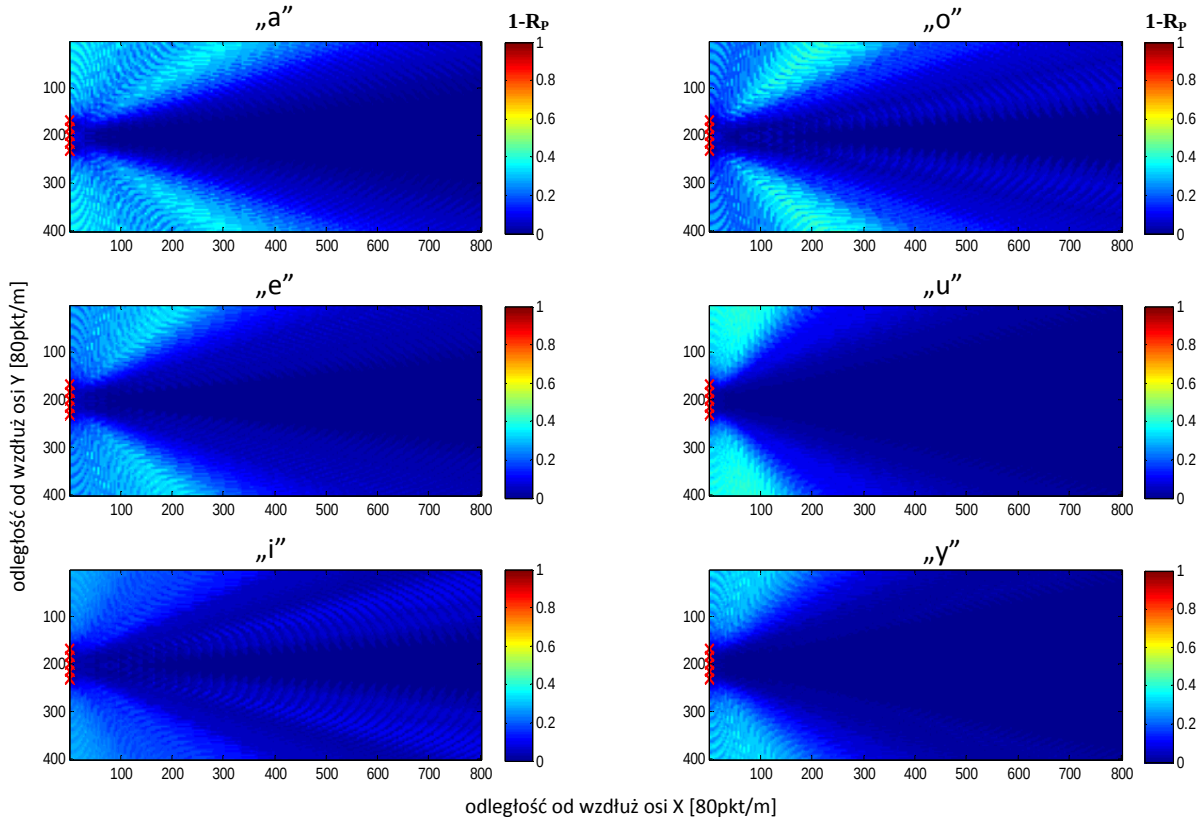
sąsiednich punktów w obszarze odsłuchu, odpowiadających odległości międzyusznej (0.2 m). Jak widać następuje tu silne odchylenie częstotliwości pierwszego formatu F1 od położenia pierwotnego, oraz zróżnicowanie pomiędzy lewym a prawym uchem.



Rysunek 6.6 Pseudowidma LPC10 sygnału (głoska „e") dla układu typu kolumnowego (5 źródeł w jednej linii, rozstaw 0.2 m - źródła idealne), w dwóch sąsiednich punktach odsłuchu odległych od siebie o 0.2 m (odległość międzyuszna). Zielony wykres odpowiada sygnałowi oryginalnemu, wykres czerwony – lewe ucho; wykres niebieski – prawe ucho.

Na rysunku 6.7 pokazano zmiany współczynnika korelacji Pearsona R_P, liczonego dla pseudowidm LPC10, dla takiego samego układu źródeł jak przy wyznaczaniu wskaźników odległości zaprezentowanych na rys. 6.2 i 6.3 (układu szyku źródeł w jednej linii). W celach porównawczych względem wskaźnika IS wartości współczynnika korelacji zostały podane w odwróconej skali (zeru odpowiada maksymalna wartość współczynnika, jedynce odpowiada wartość minimalna). Niemal we wszystkich punktach pomiarowych wartość współczynnika korelacji była stosunkowo wysoka (powyżej 0.5 w skali standardowej).

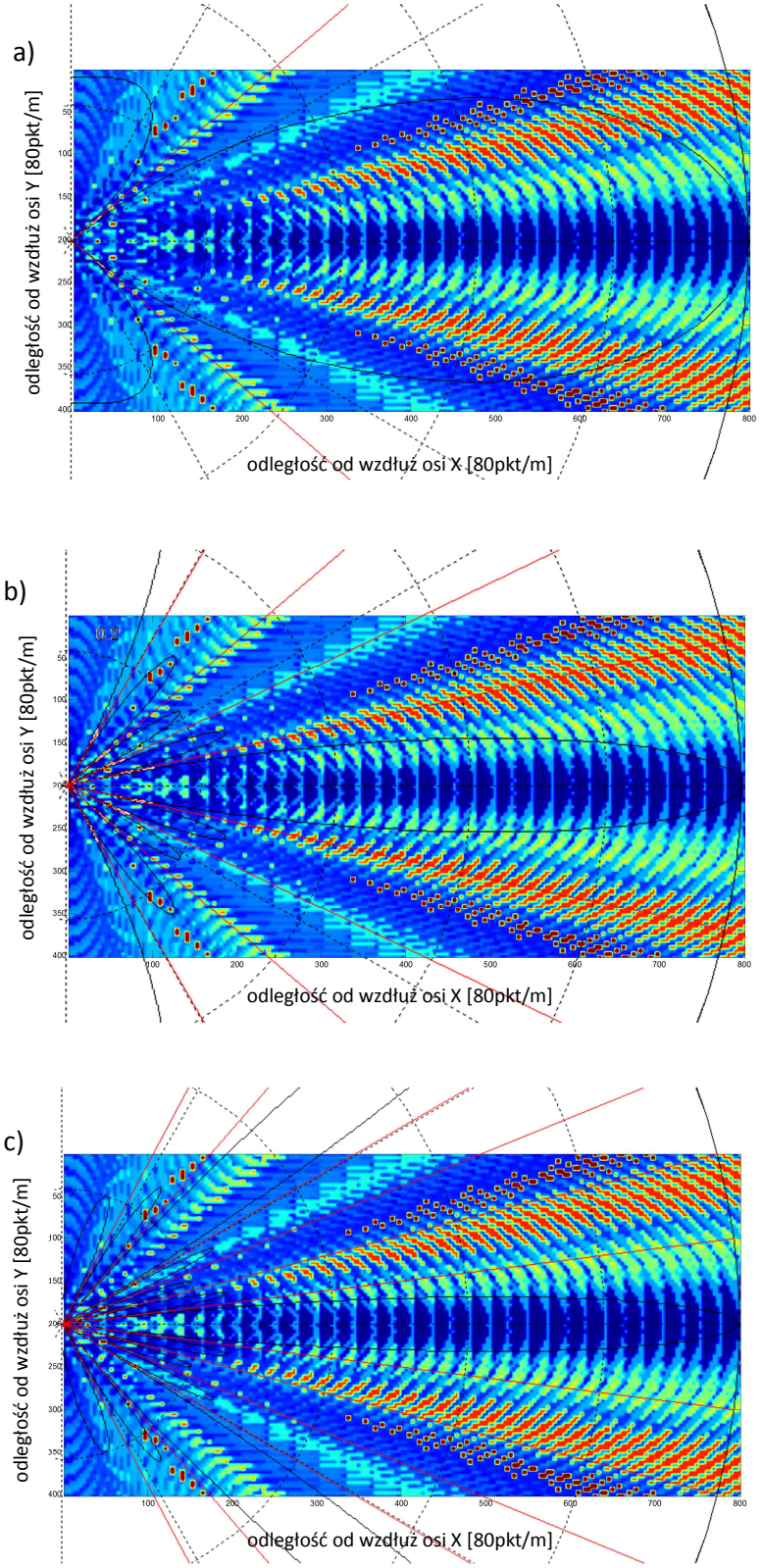
Porównanie rozkładu zmian wartości wskaźników odległości oraz współczynnika korelacji potwierdza, iż dzięki zastosowaniu wskaźników odległości otrzymano znacznie precyzyjniejszy obraz zmian charakterystyk sygnału w badanym obszarze.



Rysunek 6.7 Mapy zmienności współczynnika korelacji Pearsona, w skali odwróconej (1-R_P) gdzie 1 oznacza brak korelacji, dla sześciu samogłosek, dla liniowego układu 5 źródeł typu szyk w jednej linii, rozstaw źródeł 0.2 m (źródła idealne). Badane fonemy, układ źródeł oraz położenie płaszczyzny odsłuchu są identyczne jak na rys. 6.2 i 6.3.

W układzie typu szyk w jednej linii występuje duża zmienność wartości współczynnika IS w całym analizowanym obszarze, dla każdej badanej głoski. Obszary o dużej wartości wskaźnika IS (powyżej 1) układają się w charakterystyczne prążki, które zmieniają się wraz ze zmianą kąta odchylenia punktu odsłuchu od osi głównej apertury. Zmiany te wykazują związek z charakterystykami kierunkowymi dla danego układu źródeł, wyznaczonymi dla częstotliwości formantowych poszczególnych głosek.

Na rysunku 6.8 pokazano zależność położenia obszarów o wysokich wartościach współczynnika IS względem charakterystyk kierunkowych dla przykładowego układu źródeł. Charakterystyki kierunkowe wyznaczone zostały dla częstotliwości odpowiadających pierwszej (rys. 6.8a), drugiej (rys. 6.8b) i trzeciej (rys. 6.8c) częstotliwości formantowej dla głoski „e”. Na wykresach czerwone linie wyznaczają kierunki „zer” charakterystyk kierunkowych. Zgodnie z oczekiwaniami wzdłuż tych linii skupione są obszary o dużych poziomach zniekształceń, gdyż w naturalny sposób występuje tam filtracja poszczególnych formantów.

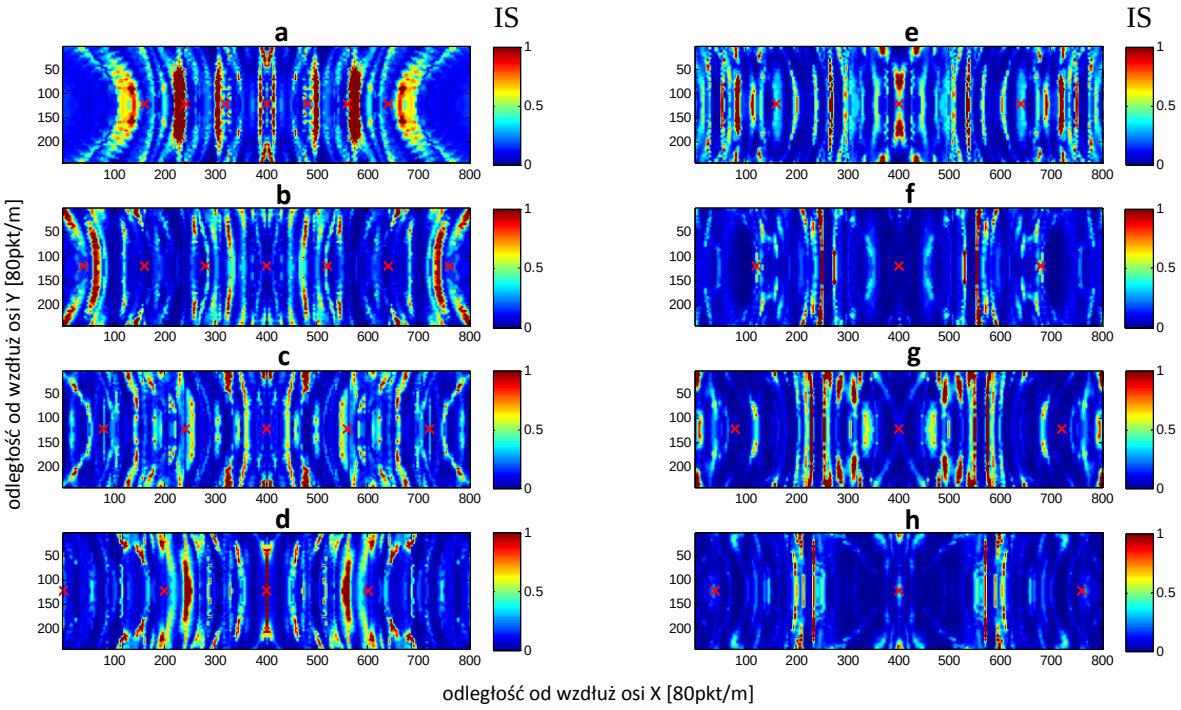


Rysunek 6.8 Wykresy charakterystyk kierunkowych, wyznaczonych dla częstotliwości odpowiadających częstotliwościom formantowym: a) pierwszego formantu F1=523, b) drugiego formantu F2=1574Hz, c) trzeciego formantu F3=2700Hz, nałożone na mapę wartości wskaźników IS, dla układu 5 źródeł rozłożonych wzdłuż linii co 0.2m (źródła idealne), dla głoski „e”. Czerwone linie wyznaczają kierunki „zer” charakterystyk kierunkowych.

6.1.2. Układ źródeł typu „ciąg komunikacyjny”

Rysunki 6.9 i 6.10 obrazują zmienność wartości współczynnika IS dla głoski „u”, dla układu źródeł określanych w dalszej części jako układ typu ciąg komunikacyjny (7 źródeł rozmieszczonych równomiernie wzdłuż linii). Płaszczyzna odsłuchu o rozmiarach (2.5m x 10m), na której wyznaczono wartości IS, znajduje się poniżej płaszczyzny źródeł o 1.5 m. Wartości wskaźników wyznaczano z rozdzielczością 80 punktów odsłuchu na 1 m, w obu kierunkach X i Y.

Dla układu typu ciąg komunikacyjny wyraźnie widoczny jest efekt polepszenia jakości akustycznej w obszarach bliskich względem pojedynczego źródła. Gdy punkt odsłuchu znajduje się blisko któregoś ze źródeł, wówczas zakłócający wpływ pozostałych źródeł jest niewielki (nieistotny), co przekłada się na bardzo małe wartości wskaźników (bliskie zeru). Obszary w pobliżu źródeł mają kolor ciemnoniebieski, oznaczający brak zniekształceń. Jest to szczególnie dobrze widoczne dla dużych odległości między źródłami (dla d > 3m). Stosunkowo silne zniekształcenia powstają natomiast w węzłach pomiędzy źródłami. Gdy odległości od punktu odsłuchu do kilku źródeł są porównywalne następuje efekt „rozmazania” wartości wskaźnika IS. Występuje duża częstotliwość przestrzenna zmian wartości wskaźnika, która zmniejsza się wraz ze wzrostem rozstawu źródeł. Na zaprezentowanej sekwencji map zmienności (rysunki od 6.9a do 6.9h) zostało pokazane jaki wpływ na rozkład wartości wskaźnika IS mają zmiany rozstawu źródeł w zakresie od w zakresie od 1m do 4.5m.

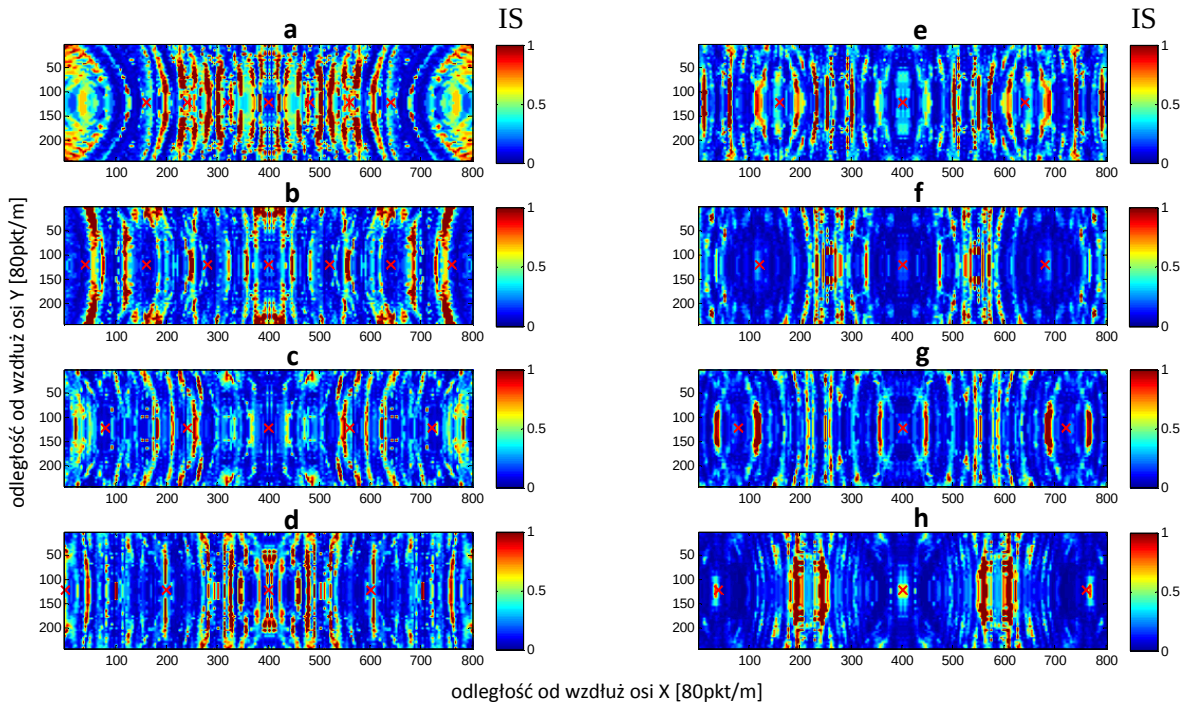


Rysunek 6.9 Mapy zmienności wskaźnika IS dla głoski „u”, dla układ typu ciąg komunikacyjny – źródła idealne. Płaszczyzna odsłuchu znajduje się 1.5 m poniżej płaszczyzny źródeł. Zmiana rozstawu źródeł w zakresie od 1m (a) do 4.5 m (h) z krokiem 0.5 m.

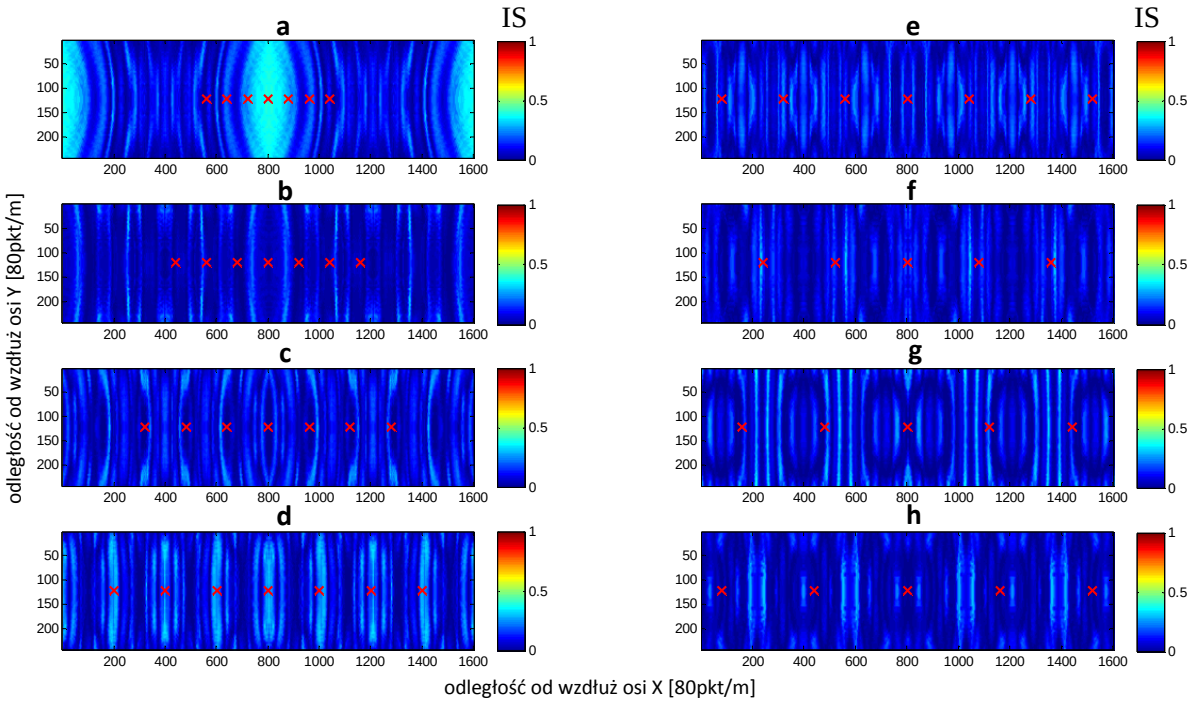
Obszary, w których zachodzą silne zjawiska interferencyjne, powodujące silne zakłócenia sygnału, odwzorowane są wysokimi poziomami wskaźników IS. Są to obszary zaznaczone w kolorystyce pomarańczowo-czerwonej, szczególnie dobrze widoczne w obszarach znajdujących się w środku pomiędzy źródłami.

W Dodatku B2 przedstawione są mapy zmienności IS dla pozostałych samogłosek polskich, dla tego samego układu źródeł typu ciąg komunikacyjny, w różnych konfiguracjach rozstawu źródeł i położenia płaszczyzny odsłuchu.

Na rysunku 6.10 pokazano mapę zmienności wskaźnika IS dla głoski „u", dla układu źródeł identycznego jak na rys. 6.9, ale obliczenia wykonano w oparciu o pomiary rzeczywistych odpowiedzi impulsowych zestawu głośników (ch-ki z rys. 3.6 i 3.7).



Rysunek 6.10 Mapy zmienności wskaźnika IS dla głoski „u", płaszczyzna odsłuchu znajduje się 1.5 m poniżej płaszczyzny źródeł, układ źródeł typu ciąg komunikacyjny. Badany fonem, układy źródeł oraz położenie płaszczyzny odsłuchu są identyczne jak na rys. 6.9, ale obliczenia wykonano dla odpowiedzi impulsowych rzeczywistych źródeł (głośników), których charakterystyki pokazane są na rys. 3.6 i rys. 3.7.



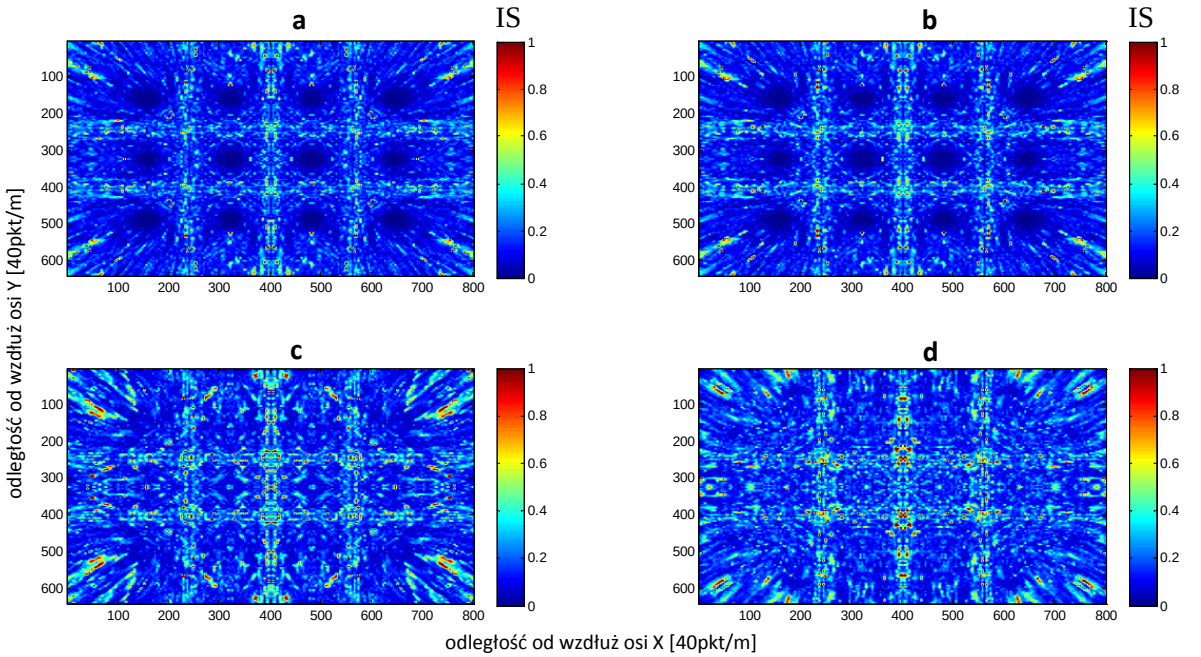
Rysunek 6.11 Mapy zmienności współczynnika korelacji Pearsona, w skali odwróconej ($1-R_P$) gdzie 1 oznacza brak korelacji, dla głoski „u", płaszczyzna odsłuchu znajduje się 1.5 m poniżej płaszczyzny źródeł, układ źródeł typu ciąg komunikacyjny (źródła idealne). Badany fonem, układy źródeł oraz położenie płaszczyzny odsłuchu są identyczne jak na rys. 6.9 i 6.10.

Rysunek 6.11 pokazuje zmiany współczynnika korelacji Pearsona liczonego dla pseudowidm LPC10, dla takiego samego układu źródeł jak przy wyznaczaniu wskaźników odległości dla układu ciągu komunikacyjnego zobrazowanych na rys. 6.9 i 6.10. Również w tym przypadku, podobnie jak dla szyku źródeł, aby ułatwić bezpośrednie porównanie względem wskaźnika IS, wartości współczynnika korelacji zostały podane w odwróconej skali (zeru odpowiada maksymalna wartość współczynnika, jedynce odpowiada wartość minimalna). Także dla tego układu uzyskano w niemal wszystkich punktach pomiarowych wartość współczynnika korelacji powyżej 0.5 (w skali standardowej). Wynik ten potwierdza wcześniejszy wniosek, iż dzięki zastosowaniu wskaźników odległości otrzymano bardziej precyzyjny obraz zmian charakterystyk sygnału w badanym obszarze.

6.1.3. Układ źródeł typu „sala audytoryjna"

Rysunek 6.12 obrazuje zmienność wartości współczynnika IS dla głoski „e", dla układu matrycowego źródeł, określanych w dalszej części jako układ typu sala audytoryjna. Obszar o wymiarach podłogi 16m x 20m nagłaśniany jest przez 12 źródeł rozmieszczonych w jednej płaszczyźnie (np. w suficie). Płaszczyzna odsłuchu, na której wyznaczono wartości IS, znajduje się

poniżej płaszczyzny źródeł w zakresie od 0.1m do 2m. Wartości wskaźników wyznaczano z rozdzielczością 40 punktów odsłuchu na 1m w obu kierunkach X i Y. Mapy zmienności wskaźnika CD (Cepstrum Distance) dla identycznego układu zaprezentowane zostały w Dodatku B3.



Rysunek 6.12 Mapy zmienności wskaźnika odległości IS (Itakura-Saito) w funkcji położenia punktu odsłuchu. Obszar o wymiarach podłogi 16m x 20m nagłaśniany jest przez 12 źródeł rozmieszczonych w jednej płaszczyźnie (np. w suficie). Płaszczyzna odsłuchu znajduje się poniżej płaszczyzny źródeł o: a) 0.1 m; b) 0.2m; c) 1m; d) 2m. W celach porównawczych wartości wskaźnika IS dla wszystkich przypadków zostały ograniczone do jednakowego zakresu [0, 1].

W tabelach 6.1 i 6.2 zestawiono porównanie wartości wskaźników STI oraz IS dla rzeczywistych odpowiedzi impulsowych, pomierzonych w sali audytoryjnej z nagłośnieniem podwieszonym w suficie (12 źródeł). Badania metodą korelacyjną przeprowadzono dla 13 punktów pomiarowych, rozmieszczonych w nagłaśnianym obszarze typu sala audytoryjna (rozdział 3.2). Wartości zamieszczone w tabeli 6.1 obliczono dla całkowitych przebiegów mierzonych odpowiedzi impulsowych, uwzględniających pełny pogłos. W tabeli 6.2 zestawiono wartości wskaźników obliczone tylko dla początkowych fragmentów odpowiedzi impulsowych (ok.20 ms) pomierzonych zgodnie z metodologią opisaną w rozdziale 3.2, odpowiadających czasom dojścia do punktu odsłuchu impulsów bezpośrednio od źródeł.

Porównanie obu zestawień potwierdza brak czułości wskaźnika STI na zniekształcenia transmitowanego przekazu, wynikające z interferencji szerokopasmowej, w układach wieloźródłowych.

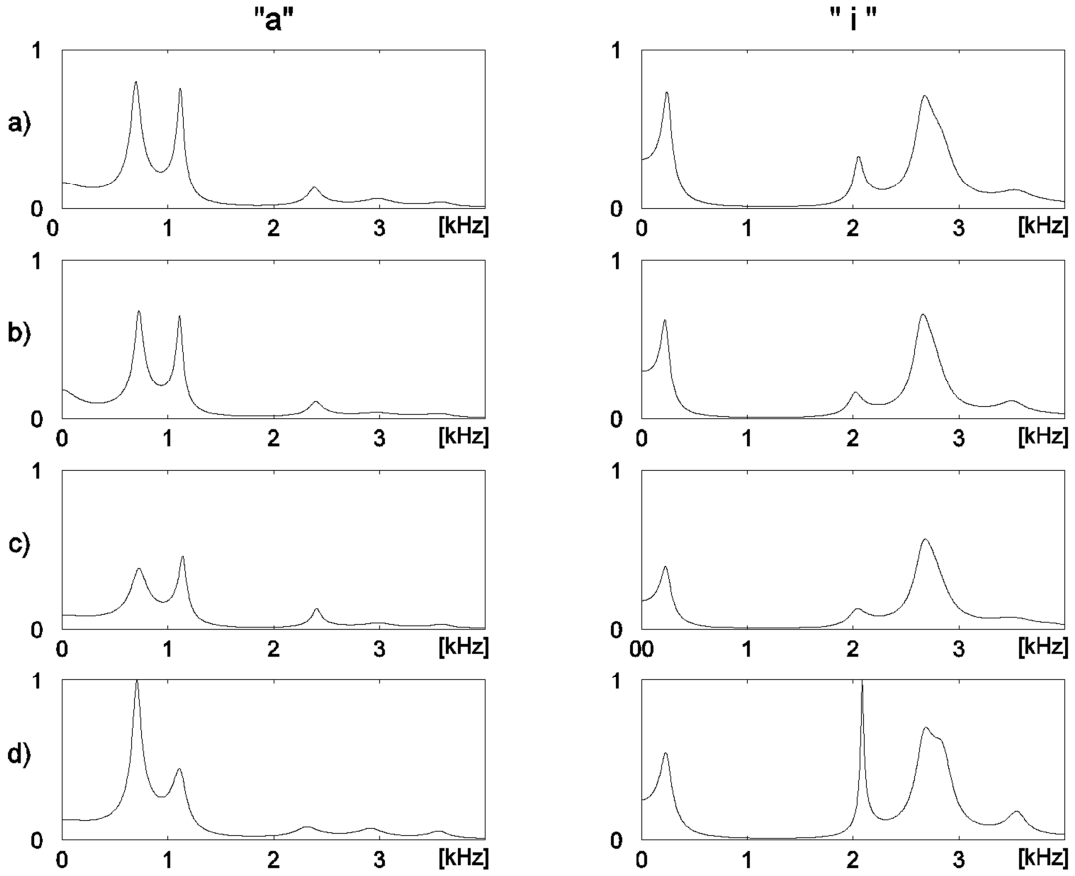
Tabela 6.1 Porównanie wartości wskaźnika STI z wartościami wskaźnika IS, dla 6 samogłosek polskich, w 12 punktach pomiarowych audytorium z nagłośnieniem umieszczonym w suficie (12 źródeł), wyznaczonych dla pełnych odpowiedzi impulsowych (ok. 1s).

		Wskaźnik IS					
pomiar	STI	„o”	„e”	„i”	„o”	„u”	„y”
IR1	0.8000	0.4482	2.6941	0.8195	0.3074	0.8341	1.2698
IR2	0.7778	0.7375	1.1625	0.2648	0.6061	0.3403	0.4202
IR3	0.7626	0.2410	1.5579	0.4121	0.7986	0.1211	1.0599
IR4	0.7498	0.3930	0.6076	0.1230	0.6180	0.2747	0.4288
IR5	0.7531	0.3248	3.5919	0.3430	0.8655	2.6133	0.4502
IR6	0.7539	0.5237	3.8202	0.5293	0.2940	2.9881	0.4251
IR7	0.7626	0.9065	0.8895	0.3393	0.1802	4.1485	0.2626
IR8	0.7642	1.4133	0.4973	0.5409	0.2887	0.1373	0.7170
IR9	0.7195	0.1536	5.6838	0.0337	0.4270	0.0904	1.7674
IR10	0.7187	0.3316	2.8455	0.2665	0.1100	0.3013	0.2224
IR11	0.7118	0.7272	1.7838	1.0471	0.2099	0.6526	0.5423
IR12	0.7362	0.3202	1.4712	0.8918	0.3112	3.2108	0.6945

Tabela 6.2 Porównanie wartości wskaźnika STI z wartościami wskaźnika IS, dla 6 samogłosek polskich, w 12 punktach pomiarowych audytorium z nagłośnieniem umieszczonym w suficie (12 źródeł), wyznaczonych dla początkowych fragmentów odpowiedzi impulsowych (ok. 20 ms), odpowiadających czasom dojścia do punktu odsłuchu impulsów bezpośrednich od źródeł.

		Wskaźnik IS					
pomiar	STI	„o”	„e”	„i”	„o”	„u”	„y”
IR1pocz	0.9951	0.6270	0.9421	0.3612	0.3059	0.1324	0.5412
IR2pocz	0.9970	0.0771	0.6804	0.3061	0.1570	0.4002	1.0190
IR3pocz	0.9965	0.3341	0.4861	0.1922	0.2976	0.4357	0.4316
IR4pocz	0.9948	0.2197	0.4553	0.2002	0.1642	0.0328	0.1197
IR5pocz	0.9957	0.3034	0.3600	0.4259	0.1793	0.1691	0.7824
IR6pocz	0.9971	0.4142	0.3711	0.3137	0.4599	0.1389	1.0956
IR7pocz	0.9961	0.3249	0.4534	0.2238	0.1626	0.1296	0.5784
IR8pocz	0.9954	0.2564	0.9267	0.5767	0.1975	0.1103	0.7199
IR9pocz	0.9956	0.0905	0.2317	0.0885	0.2484	0.1664	0.2314
IR10pocz	0.9938	0.3101	0.1637	0.2139	0.2323	0.3220	0.1405
IR11pocz	0.9962	0.9464	0.7724	0.9092	0.2954	0.3698	1.0908
IR12pocz	0.9970	0.4154	0.9901	0.4779	0.6538	0.2037	1.3698

Jak pokazano, zalecane przez normy obliczenia indeksów zrozumiałości STI, bazujące na pomiarach tzw. funkcji przenoszenia modulacji MTF, zaprojektowane są głównie dla oceny wpływu na jakość przekazu zjawisk pogłosowych. Nie zawsze w pełni odzwierciedlają one wszystkie aspekty zmniejszenia zrozumiałości, gdyż wiązek funkcji MTF ze zrozumiałością jest w istocie pośredni [60] [65] [11]. STI, które dobrze charakteryzuje warunki pogłosowe, nie uwzględnia niemal w ogóle zjawisk związanych z interferencją szerokopasmową.



Rysunek 6.13 Przykład zmian pseudowidma LPC10 sygnałów głosek „a” oraz „i” po przejściu przez modelowany system w wybranym punkcie odsłuchu dla: a) 1 źródło, b) 2 źródła, c) 4 źródła, d) 12 źródeł; logarytmiczna skala częstotliwości.

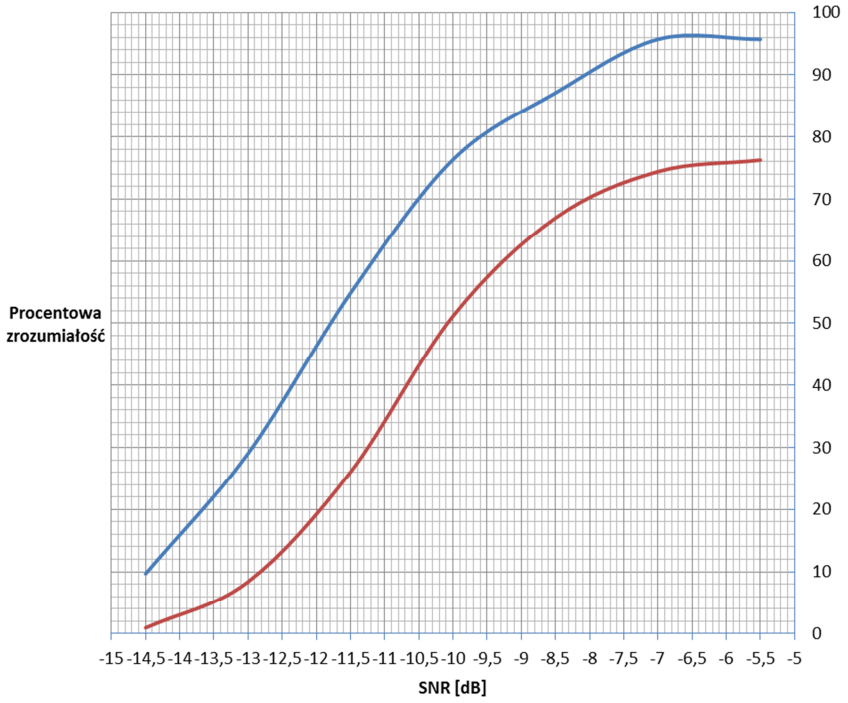
Wyniki obliczeń symulacyjnych wskazują, że wzrost liczby współbieżnie działających źródeł sygnału, powoduje powstawanie silnych zniekształceń formantów (rys. 6.13). Przy wzroście liczby źródeł, od pojedynczego, poprzez 2 i 4 źródła aż do układu 12-tu źródeł, następuje wyraźna deformacja proporcji między poszczególnymi formantami.

6.2. Badanie subiektywne degradacji jakości przekazu w systemach wielośródłowych

W [26] opisano badania przeprowadzone na Wydziale ETI PG, pod kierownictwem dra hab. inż. Henryka Lasoty, dotyczące wpływu interferencji szerokopasmowej na zrozumiałość przekazu głosowego. Przeprowadzono test rozpoznawalności polskich trypletów cyfrowych w systemie

wielośródłowym. Badania przeprowadzono w formie komputerowego testu internetowego, w którym udział wzięło 51 słuchaczy w wieku od 18 do 60 lat.

Wyniki przeprowadzonych testów ilustruje rys. 6.14, przedstawiający krzywe zrozumiałości dla mowy na tle szumu (linia niebieska) oraz dla mowy z wprowadzonymi zniekształceniami występującymi w układach wielośródłowych (linia czerwona). Wyraźnie zauważalne jest przesunięcie krzywej zrozumiałości dla mowy zniekształconej w kierunku niższego poziomu szumu. Oznacza to, iż ten sam procent poprawnych odpowiedzi uzyskiwano we wszystkich badaniach dla niższego poziomu szumu. Aby uzyskać 50% zrozumiałość poziom szum musiał być o ok. 1.7 dB niższy w przypadku mowy bez zniekształceń.

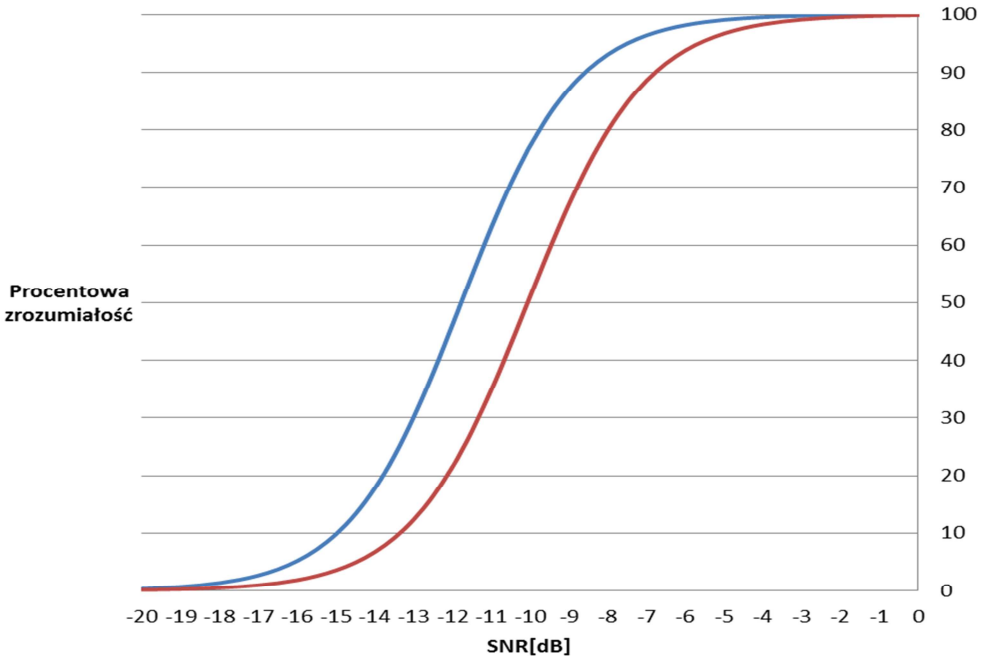


Rysunek 6.14 Krzywe zrozumiałości: niebieska linia - mowa na tle szumu; czerwona - mowa dodatkowo zniekształcona [26].

Na podstawie wyznaczonych krzywych zrozumiałości określono próg percepcji mowy (SRT) oraz wyznaczono nachylenie ***s*** wykresów zrozumiałości, niezbędne do aproksymacji funkcji zrozumiałości *p* zgodnie z zależnością (5.18). Nachylenie zbocza funkcji ***s***, wyznaczane jest w punkcie SRT w jednostce [%/dB]. Obie wartości zostały wyznaczone przy założeniu jednorodnego nachylenia zbocza funkcji w punkcie 50% zrozumiałości mowy, za pomocą proporcji matematycznej. Dla przedstawionych powyżej krzywych, oba parametry przedstawiają się następująco (tabela 6.3):

Tabela 6.3 Parametry funkcji zrozumiałości mowy [26].

Metoda badania	SNR [dB]	s [%/dB]
Wyłącznie szumy	-11,8	17,2
Szumy i zniekształcenia	-10,1	16,7



Rysunek 6.15 Aproksymowane funkcje zrozumiałości: niebieska linia - mowa na tle szumu, czerwona - mowa dodatkowo zniekształcona [26].

Obie krzywe zrozumiałości mowy (rys. 5.15) charakteryzują się podobną wartością nachylenia zboczy funkcji co świadczy o dobrej porównywalności wyników.

Opisane badania potwierdzają niekorzystny wpływy interferencji szerokopasmowej na zrozumiałość mowy. Efekt ten jest szczególnie silnie zauważalny w warunkach występowania wysokiego poziomu szumu otoczenia.

Wykorzystywana w opisywanych badaniach aplikacja służąca do oceny stopnia zniekształcenia sygnału mowy, zrealizowana w formie komputerowego testu internetowego, może zdaniem autora zostać rozszerzona o mechanizmy zawarte w aplikacji opisanej w rozdziale 3.3. Pozwoliłoby to na bieżące weryfikowanie (a nawet kalibrowanie) systemu na podstawie wyznaczonych parametrów oraz dałoby możliwość testowania jakości sygnału po przejściu przez kodek LPC10.

7. PODSUMOWANIE

Celem głównym niniejszej pracy była weryfikacja tez, iż superpozycja sygnałów pochodzących ze współbieżnych źródeł rozmieszczonych w różnych odległościach od punktu odsłuchu wywołuje efekt zniekształcenia parametrów czasowo-częstotliwościowych sygnałów szerokopasmowych oraz, że wskazane jest zastosowanie obiektywnej, powtarzalnej miary degradującego wpływu szerokopasmowej interferencji liniowej na zrozumiałość sygnałów mowy.

Wymagało to przeprowadzenie analizy czasowo-częstotliwościowej układów akustycznych przy wykorzystaniu metody odpowiedzi impulsowych oraz opisanie charakteru zniekształceń powstających na skutek przestrzennego rozkładu źródeł akustycznych. Przeprowadzono również szereg badań symulacyjno-pomiarowych obrazujących wpływ interferencji szerokopasmowej na parametry sygnałów mowy, w wyniku czego zaproponowano zastosowanie obiektywnych wskaźników miar stopnia ich zniekształcenia.

W ramach samodzielnie przeprowadzonych badań, realizując postawione cele, autor zrealizował następujące zadania.

1. Wykazał, że stosowanie opisu pola akustycznego zespołów źródeł szerokopasmowych w kategoriach systemów LTI, pozwala na identyfikację i analizę niekorzystnych zjawisk wynikających z interferencji szerokopasmowej, których to efekty nie są widoczne przy zastosowaniu powszechnie przyjmowanego modelu natężeniowego (optycznego).
2. Przeprowadził pomiary rzeczywistego wielоźródłowego systemu szerokopasmowego w sali audytoryjnej. Na podstawie pomiarów odpowiedzi impulsowych rzeczywistych źródeł dźwięku (głośników) oraz precyzyjnych pomiarów położenia względem punktu odsłuchu (dalmierz laserowy) zasymulował rzeczywiste odpowiedzi impulsowe, uniezależnione od zakłóceń pochodzących od odbić. W celach testowych pominięto kierunkowość poszczególnych źródeł.
3. Skonstruował model symulacyjny zespołów źródeł szerokopasmowych wykorzystujący metodę odpowiedzi impulsowych do wyznaczenia wskaźników zmian parametrów predykcyjnych i cepstralnych sygnałów głosowych. Wszystkie badania przeprowadził używając programów zaprojektowanych i zaimplementowanych przez siebie w środowisku Matlab. Opracował oryginalne narzędzie obliczeniowe do analizy własności wielоźródłowych systemów akustycznych, w oparciu matematyczne modele miar odległości pomiędzy wektorami parametrów.
4. Zaproponował efektywną metodę oceny stopnia zniekształceń pola akustycznego poprzez badanie zmienności wskaźników odległości pomiędzy wektorami obiektywnych parametrów sygnałów mowy, takich jak współczynniki predykcyjne LPC czy współczynniki MFCC. Wybrał cztery miary wskaźników odległości LLR (Log-Likelihood

Ratio), IS (Itakura-Saito), CD (Cepstrum Distance), melCD (MFCC Distance), które służą do oceny stopnia zmian jakości pola akustycznego w obszarach badanych układów wieloźródłowych. Metoda ta umożliwia:

- wykonanie licznych eksperymentów laboratoryjnych,
- gromadzenie danych z modelu fizycznego,
- bieżącą wizualizację wyników z dużą rozdzielczością przestrzenną (0.01m),
- optymalizację procedur projektowych.

Do zalet i korzyści, wynikających z takiej metody laboratoryjnej zaliczyć należy:

- kształtowanie intuicji projektowej,
- minimalizację kosztów organizacyjnych testów eksperymentalnych.

5. Wykazał dużo większą zmienność przestrzenną wybranych wskaźników oceny jakości pola niż wynika to z bezpośredniego obliczenia współczynników korelacji Pearsona pomiędzy sygnałami: oryginalnym i zniekształconym, co potwierdza dużą czułość takiej metody oraz daje szerokie możliwości zastosowania do precyzyjnych badań jakości pola (np. subtelnych zmian sygnału związanych z odległością międzyuszną).
6. Zgromadził znaczna bazę zbiorów odpowiedzi impulsowych, zarówno modelowych układów akustycznych jak i rzeczywistych systemów wieloźródłowych, na podstawie których wyekstrahował odpowiedzi uniezależnione od odbić. Pomiary zrealizował metodą korelacyjną, z wykorzystaniem sekwencji MLS.
7. Przeprowadził szereg testów symulacyjnych i eksperymentalnych, w tym:
    - testy modelu równomiernego szyku liniowego ze zmiennym rozstawem źródeł,
    - testy modelu układu ciągu komunikacyjnym ze zmiennym rozstawem źródeł i wysokości nad płaszczyzną odsłuchu,
    - testy wybranych modeli układów z nierównomiernym rozkładem źródeł.

W oparciu o wyniki badań eksperymentalnych i symulacyjnych udowodniono postawione tezy. Pokazano, iż w układach akustycznych złożonych z wielu współbieżnie promieniujących źródeł szerokopasmowych, w wyniku superpozycji fal, występuje zjawisko interferencji szerokopasmowej, przekładającej się na zniekształcenia funkcji przenoszenia układu. Zniekształcenia te są inne w każdym puncie nagłaśnianego obszaru.

Wykazano występowanie rzutowania zniekształceń powstałych w polu takich układów na zmiany charakterystycznych parametrów sygnałów głosowych, co szczególnie uwidacznia się w zmianach położenia i poziomów formantów. Efekt modyfikacji własności transmitowanej mowy można porównać do zniekształcenia funkcji przenoszenia filtru traktu głosowego (np. wynikającego z choroby narządu głosowego). Wykazano możliwość zastosowania, do oceny degradującego wpływu szerokopasmowej interferencji liniowej na zrozumiałość sygnałów mowy,

miar opartych na zmianach obiektywnych parametrów związanych z mechanizmem generowania głosu, a szczególnie parametrów predykcji liniowej (LPC).

Wyniki przeprowadzonych badań są kluczowe dla zaprojektowania uzupełnienia standardów opisujących pomiary zrozumiałości przekazu słownego w systemach dźwiękowych instalowanych w miejscach publicznych (PAS) oraz standardów badania poziomu zniekształceń w dźwiękowych systemach ostrzegawczych (DSO).

Bibliografia

[1] P. M. Morse, Vibration and Sound, New York, Toronto, Londyn: McGraw-Hill Book Company, Inc., 1948.

[2] I. Małecki, Teoria fal i układów akustycznych, Warszawa: PWN, 1964.

[3] A. D. Pierce, Acoustics, An Introductionto Its Phisical Principles and Applications, New York: Acoustical Society of America, 1989.

[4] B. D. Steinberg, Principles of Aperture and Array System Design, New York, London, Sydney, Toronto: John Wiley and Sons, 1976.

[5] R. Makarewicz, Dzwięki i fale, Poznań: Wydawnictwo naukowe UAM, 2004.

[6] H. Lasota, R. Salamon and B. Delannoy, "Acoustic diffraction analysis by the impulse response method.," *J.Acoust.Soc.Am,* vol. 76, pp. 280-290, 1984.

[7] W. Rdzanek, Teoria pola akustycznego, Rzeszów: Wydawnictwa WSP, 1982.

[8] *PN-EN 60849:2001. Dźwiękowe systemy ostrzegawcze..*

[9] A. C. Gade, „Part C Architectual Acoustics, Acoustics in Halls Speech and Music," w *Handbook of Acoustics*, New York, Springer, 2007, pp. 301-350.

[10] S. R. Quackenbush, T. P. Barnwell III i M. A. Clements, Objective Measures of Speech Quality, New Jersey: Prentice-Hall Inc., 1988.

[11] *IEC 60268-16 Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index,* 2003-05.

[12] T. Houtgast and H. Steeneken, "Evaluation of Speech Transmission Channels by using Artificial Signals," *Acustica,* vol. 25, p. 355–367, 1971.

[13] H. Lasota, „Kierunkowość dla pobudzeń impulsowych," w *Mat. IV Sympozjum Hydroakustyki, s. 155-162*, Jastrzębia Góra, 1987.

[14] H. Lasota, R. Mazurek i M. Miler, „Pole akustyczne źródeł szerokopasmowych i ich zespołów," w *Mat. XLIX OSA, s. 589 – 594*, Warszawa, 2002.

[15] P. R. Stepanishen, "Transient radiation from pistons in an infinite baffle," *J.Acoust.Soc.Am,* vol. 49, pp. 1629-1638, 1971.

[16] H. Lasota i R. Mazurek, „Modelowanie i pomiary nagłośnienia audytorium, Pomiary Automatyka Kontrola, nr 3, str. 148-152, 2008".

[17] H. Lasota i R. Mazurek, „Interferencja szerokopasmowa w wieloźródłowych systemach akustycznych, Zeszyty naukowe wydziału elektroniki, telekomunikacji i informatyki Politechniki Gdańskiej 2008, str. 495-500".

[18] H. Lasota i R. Mazurek, „Broadband interference in speech reinforcement systems, Proceedings of the 1st International Conference on Information Technology Gdańsk, 19-21 May 2008, str. 329-332".

[19] *ISO 3382, Acoustics - Measurement of the reverberation time of rooms with reference to other acoustical parameter,* 1997.

[20] A. Dobrucki, Przetworniki elektroakustyczne, Warszawa: Wydawnictwa Naukowo-Techniczne, 2007.

[21] R. Mazurek i H. Lasota, „Application of maximum length sequences to impulse response measurement of hydroacoustic communications systems," *Hydroacoustics,* tom 10, p. 123–130 , 2007.

[22] D. Riffe i J. Vanderkooy, „Transfer-Function Measurement with Maximum-Length Sequences," *Journal of the Audio Engineering Society,* tom 37, nr 6, pp. 419-444, 1989.

[23] R. Mazurek i H. Lasota, „Application of Maximum-Length Sequences To Impulse Response Measurement Of Hydroacoustic Communications Systems," *Hydroacoustics,* pp. 123-130, 2007.

[24] M. Cohn i A. Lempel, „On Fast M-Sequence Transforms," *IEEE Transactions on Information Theory,* tom 23, nr 1, pp. 135-137, 1977.

[25] T. P. Zieliński, Cyfrowe przetwarzania sygnałów. Od teorii do zastosowań, Warszawa: Wydawnictwa Komunikacji i Łączności, 2005.

[26] J. Lach, „Wpływ rozmieszczenia i liczby źródeł dźwięku, na jakość przekazu słownego," Praca magisterska Politechnika Gdańska WETI, Gdańsk, 2011.

[27] B. Kostek, Perception-Based Data Processing in Acoustics, Application to Musuc Information Retrieval and Psychophysiology of Hearing, Warszawa: Springer, 2005.

[28] „Podstawowe wiadomości na temat sygnału mowy i traktu głosowego," [Online]. Available: http://sound.eti.pg.gda.pl.

[29] The National Center for Voice & Speech, [Online]. Available: www.ncvs.org.

[30] J. D. Markel and A. H. Gray, Linear prediction of speech, New York: Springer-Verlag, 1976.

[31] A. Czyżewski, Dzwięk cyfrowy. Wybrane zagadnienia teoretyczne, technologia, zastosowania, Warszawa: Akademicka Oficyna Wydawnicza EXIT, 1998.

[32] J. Schur, "Uber Potenzreihm, die in Inner des Einheitskreises beschrank sind," *J. fuer die Reine and Angewandte Mathematiek,* vol. 147, pp. 205-232, 1917.

[33] M. R. Schroeder, "Direct (nonrecursive) Relation Between Cepstrun and Prediction Coefficients," *IEEE Transaction on acoustisc, speech, and signal processing,* Vols. ASSP-29, no. 2, pp. 297-301, 1981.

[34] S. Wu and L. C. W. Pols, "A distance measure for objective qality evaluation of speech communication channels uing also dynamic spectral features," *Institute of Phonetic Sciences, University of Amsterdam,* vol. Proceedings, no. 20, pp. 27-42, 1996.

[35] S. B. Davis i P. Mermelstein, „Comparison of Parametric representations for Monosyllabic Word Recognition in Continuously Spoken Sentences," *IEEE Transactions Acoustics, Speech and Signal Processing,* tom 28, nr 4, pp. 375-366, 1980.

[36] H. Hermansky, „Perceptual Linear Predictive (PLP) Analysis of Speech," *Journal Acoustical Society of America,* tom 87, nr 4, p. 1738–1752, 1989.

[37] H. Hermansky i N. Morgan, „RASTA Processing of Speech," *IEEE Transactions on Speech and Audio Processing,* tom 2, nr 4, pp. 578-589, 1994.

[38] *ANSI S3.5-1969, (R1986),Calculation of the Articulation Index,Meth,* 1969.

[39] P. Godlewski, M. J. Trzaskowska i B. Mucha, „Metody obiektywnej oceny jakości usługi głosowej QoS w sieciach łączności elektronicznej oraz urządzenia do takiej oceny i do badania dostępności "usług" poprzez numery alarmowe - etap1," Instytut Łączności, Warszawa, 2006.

[40] *IEC 60849 (Ed. 2.0 1998-02) Sound systems for emergency purposes.*

[41] *PN-90/T-05100; Analogowe łańcuchy telefoniczne - Wymagania i metody pomiaru wyrazistości logatomowej,* 1993.

[42] *ITU-T Recommendation P.800: Method for subjective determination of transmission quality,* 1996.

[43] *ITU-T P.830: Subjective Performance Assessment of Telephone-Band and Digital Codecs,* 1996.

[44] *ITU-T P.810: Modulated Noise Reference Unit (MNRU),* 1996.

[45] *ETSI ETR 250; Speech communication quality from mouth to ear for 3,1 kHz handset telephony,* 1996.

[46] *ETSI EG 201 377-1: Speech Processing, Transmission and Quality Aspects (STQ); Specification and measurement of speech transmission quality,* 1999.

[47] S. Brachmański, „Subiektywne metody oceny jakości transmisji mowy w cyfrowych kanałach

telekomunikacyjnych”.*Instytut Telekomunikacji i Akustyki, Wrocław.*

[48] K. Baściuk, S. Brachmański, W. Majewski i W. Myslecki, „Ocena jakości transmisji mowy w kanałach telekomunikacyjnych,” *Akustyka w Technice, Medycynie i Kulturze, Granty KBN 1996-1999,* pp. 141-154, 1999.

[49] *ANSI/ASA S3.5-1997 (R2012), American National Standard Methods for Calculation of the Speech Intelligibility Index,* 1997.

[50] V. M. A. Peutz, „Articulation loss of consonants as a criterion for speech transmission in a room.,” *J. Audio Eng. Soc. 19,* p. 915–919, 1971.

[51] R. Plomp i A. M. Mimpen, „Improving the reliability of testing the speech reception threshold for sentences.,” *Audiology,* tom 8, nr 52, 1979.

[52] E. Ozimek, D. Kutzner, A. Sęk i A. Wicher, „Polish sentence tests for measuring the intelligibility of speech in interfering noise,” *International Journal of Audiology,* tom 48, pp. 433-443, 2009.

[53] K. Wagener, „Report on an optimized inventory of Speech-based auditory screening & impairment tests for six languages,” D-1-9 , 2009.

[54] E. Ozimek, D. Kutzner, A. Sęk i A. Wicher, „Development and evaluation of Polish digit triplet test for auditory screening,” *Speech Communication,* tom 52, p. 307–316, 2009.

[55] M. Karjalainen, „Anew Auditory Model for the Evaluation of Sound Quality of Audio Systems,” *IEEE ICASSP,* pp. 608-611, 1985.

[56] *ITU-T P.861: Objective Quality Measurement of Telephone-Band Speech Codecs,* 1998.

[57] *ITU-T Recommendation P.862: Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs.*

[58] *ITU-R Recommendation BS 1387: Method for Objective Measurements of Perceived Audio Quality (PEAQ),* 1998.

[59] S. Brachmański i J. Zuk, „Związek miedzy wyrazistością logatomową a wskaźnikiem STI w analogowych kanałach telekomunikacyjnych dla języka polskiego,” *Instytut telekomunikacji i Akustyki, Politechnika Wrocławska.*

[60] T. Houtgast and H. Steeneken, "The Modulation Transfer Function in Rooms Acoustics as a Predictor of Speech Intelligibility," *Acoustica,* vol. 28, no. 1, pp. 66-73, 1973.

[61] P. Pruchnicki, „Metody pomiaru parametrów akustycznych pomieszcze,” w *IV Sympozjum Nowości w Technice Audio*, Wrocław, 1997.

[62] F. Itakura, „Minimum prediction residual principle applied to speech recognition,” *IEEE Transactions on Acoustics, Speech and Signal Processing,* tom 23, nr 1, pp. 67 - 72, 1975.

[63] F. Itakura, „Line spectrum representation of linear,” *J. Acoust. Soc. Am.,* tom 57, nr 537(A), 1975.

[64] W. A. Akshya K. Swain, „Estimation of LPC Parameters of Speech Signals in Noisy Environment,” *TENCON 2004. 2004 IEEE Region 10 Conference (Volume:A),* pp. 139 - 142, 2004.

[65] *PN-EN 60268-16, Urządzenia systemów elektroakustycznych. Cz.16: Obiektywna ocena zrozumiałości mowy za pomocą wskaźnika transmisji mowy,* 2005.

[66] *ITU-T Recommendation P.563, Single-ended method for objective speech quality assessment in narrow-band telephony applications,* 2004.

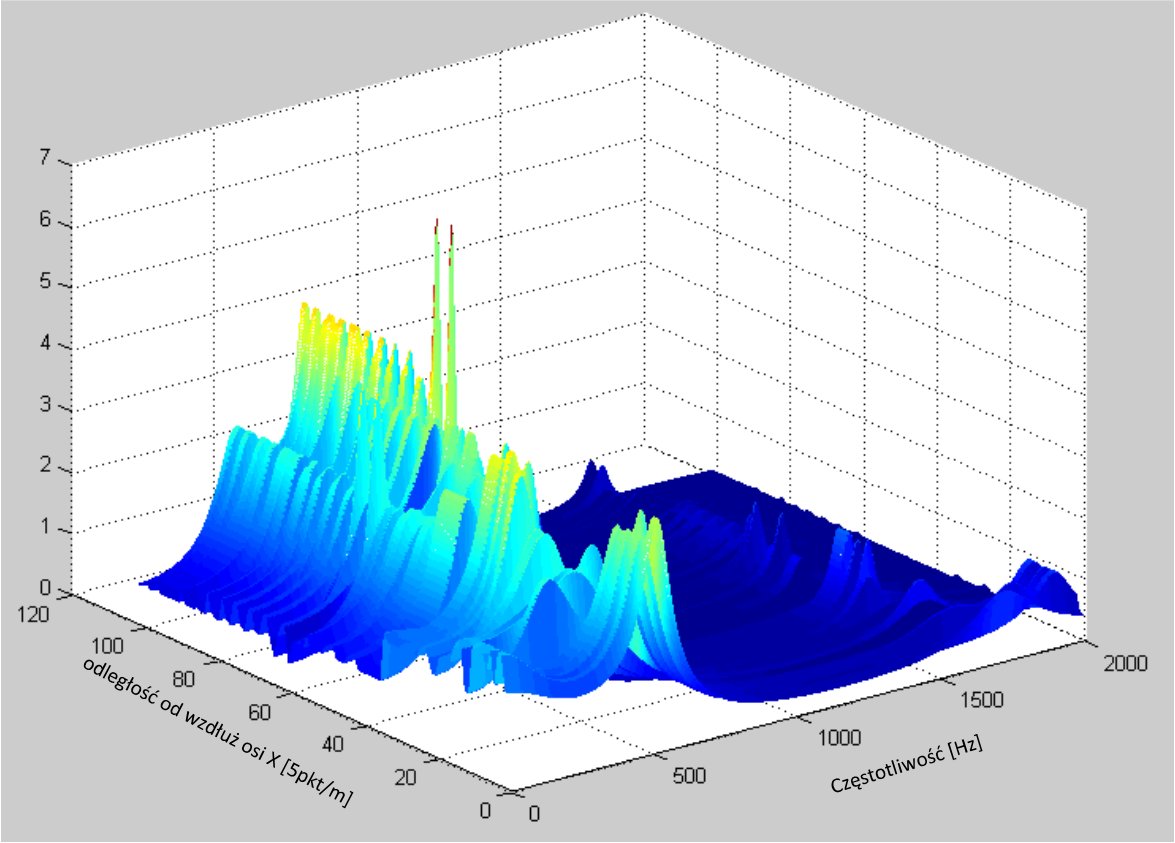
[67] R. Sinclair, "The Design of Distributed Sound Systems from Uniformity of Coverage and Other Sound-Field Con-sideration," *Journal of the AES,* vol. 30(12), p. 871–881, 1982.

[68] J. Adamczyk, H. Lasota, R. Mazurek i M. Miler, „Badanie dźwiękowego systemu ostrzegawczego w sali kinowej pod kątem zgodności z normą,” w *Materiały konferencyjne OSA'03*, Szczyrk, 2003.

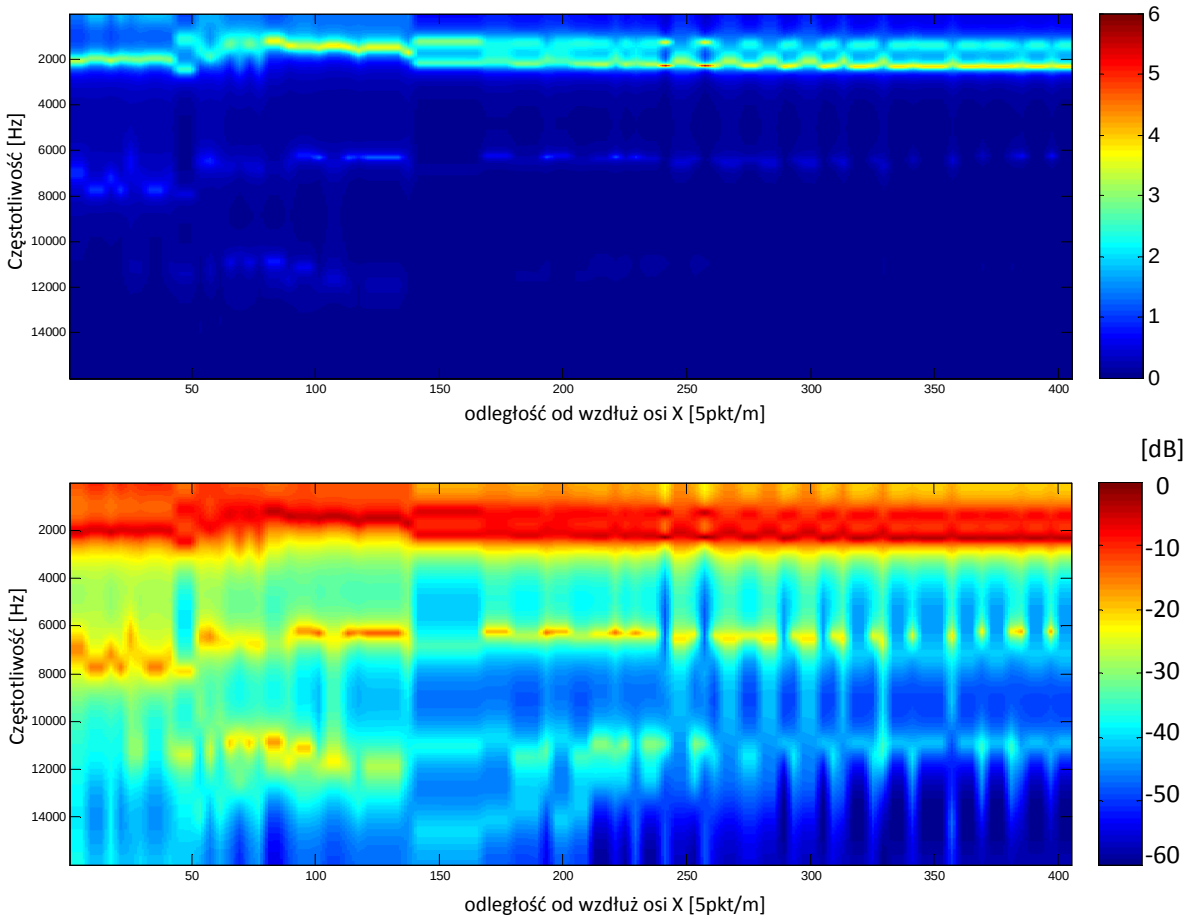
[69] J. S. Bradley, H. Sato and M. Picard, "On the importance of early reflections for speech in rooms," *Journal of the ASA,* vol. 113(6), p. 3233–3244, 2003.

DODATEK A

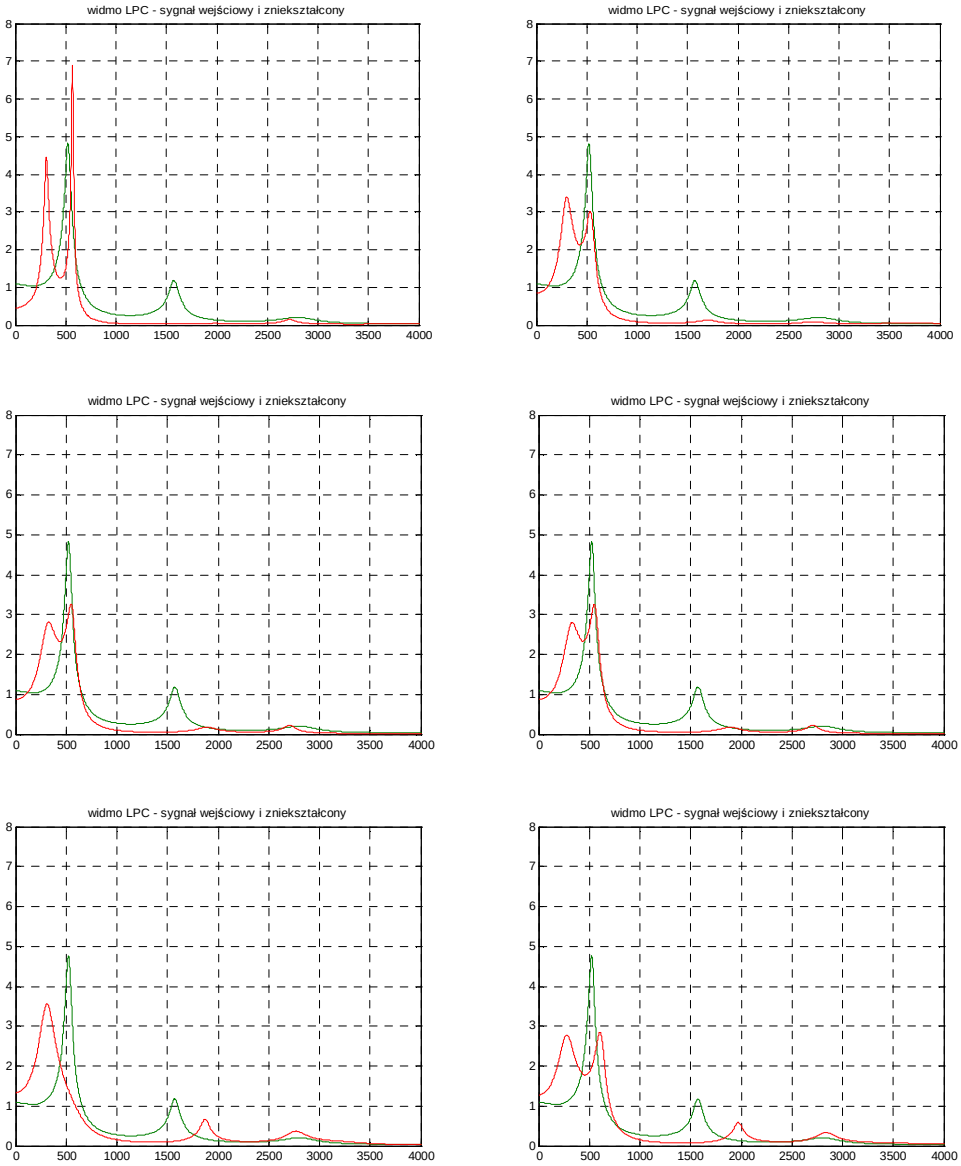
Poniżej przedstawiono zmienność przebiegu pseudowidma LPC, dla rzędu predykcji N=10, wzdłuż wybranego kierunku pola, w zobrazowaniu dwu- (rys. A.2) i trójwymiarowym (rys. A.1), oraz jego przekroje dla sześciu wybranych punktów odsłuchu (rys. A.3 i A.4). Wyraźnie widoczne są zniekształcenia kształtu pseudowidma LPC, silnie zmieniające się wraz ze zmianą położenia punktu odsłuchu.



Rysunek A.1 Zobrazowanie 3D zmienności pseudowidma LPC dla rzędu predykcji N=10, dla głoski „e”, dla układu 5 źródeł rozłożonych w linii co 0.2m. Punkt odsłuchu przesuwany wzdłuż linii równoległej do osi głównej apertury x=4.4m (znajdującej się poza pasem przyosiowym).



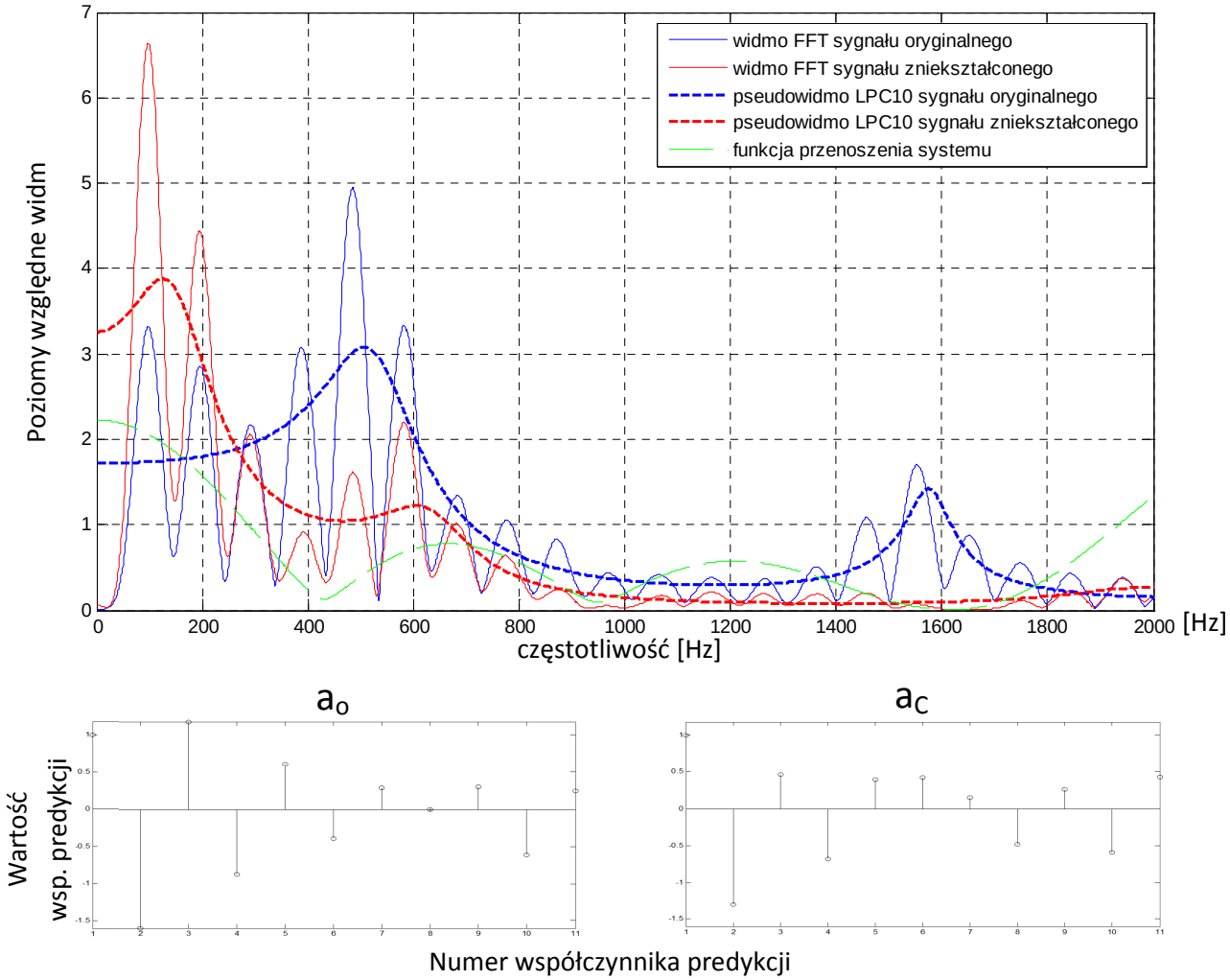
Rysunek A.2 Zmienność przebiegu pseudowidma LPC dla rzędu predykcji N=10, dla głoski „e”, dla układu 5 źródeł rozłożonych w linii co 0.2m. Punkt odsłuchu przesuwany wzdłuż linii równoległej do osi głównej apertury x=4.4m (znajdującej się poza pasem przyosiowym), a) w skali liniowej; b) w skali decybelowej.



Rysunek A.3 Przykłady zmian w kształcie pseudowidma LPC dla rzędu predykcji N=10, dla głoski „e”, dla układu 5 źródeł rozłożonych w linii co 0.2m, w wybranych punktach odsłuchu (x,y), α – kąt odchylenia od osi głównej apertury,

a) x=4.9m, y=8.0m, α = 17 st. (IS = 7.41);
b) x=4.9m, y=5.2m, α = 27 st. (IS = 1.89);
c) x=3.9m, y=3.4m, α = 22 st. (IS = 2.55);
d) x=3.4m, y=2.2m, α = 21 st. (IS = 2.50);
e) x=4.1m, y=1.8m, α = 42 st. (IS = 3.29);
f) x=4.4m, y=1.1m, α = 60 st. (IS = 6.38);

Linią zieloną oznaczono pseudowidmo sygnału oryginalnego, linią czerwoną sygnału zniekształconego, skala liniowo-liniowa.



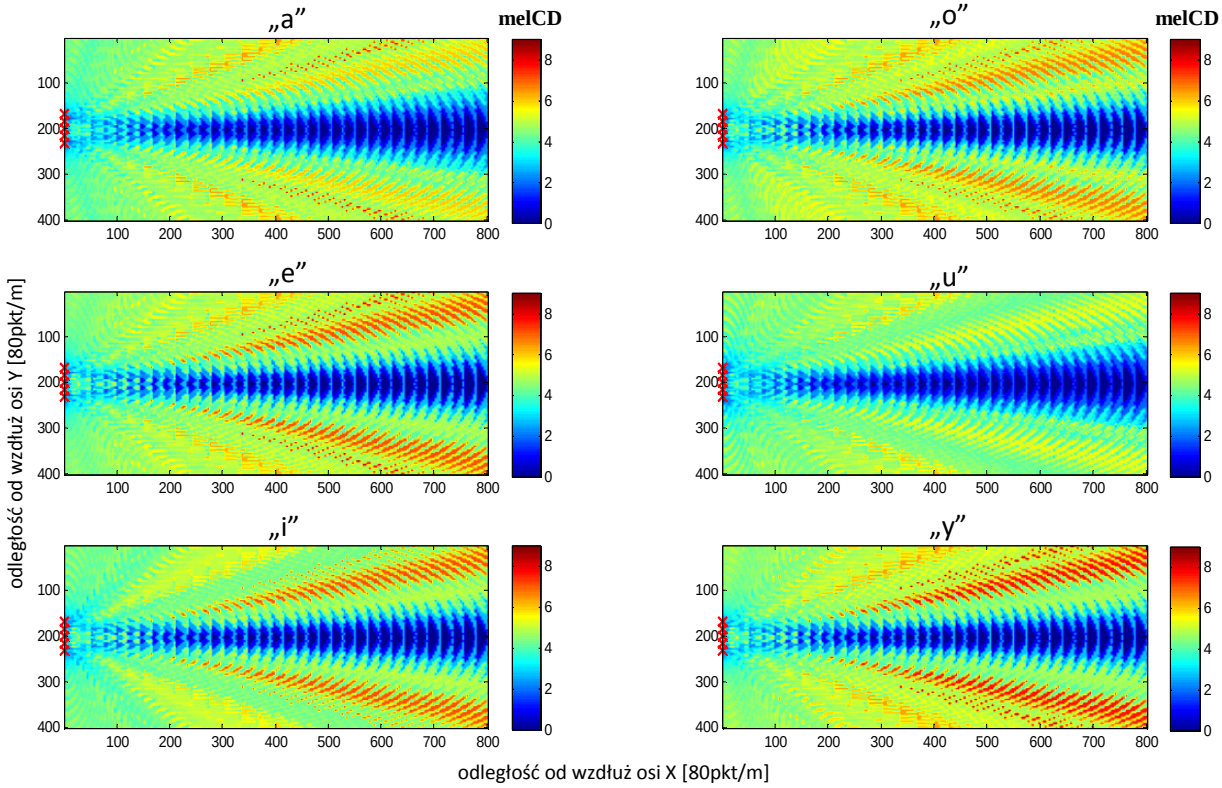
Rysunek A.4 Zniekształcenia widma FFT oraz pseudowidma LPC10 sygnału (głoska „e") dla układu typu szyk źródeł w jednej linii (5 źródeł w jednej linii, rozstaw 0.2 m - źródła idealne), w punkcie odsłuchu x=1.55 m, y=0.6 m , kąt = 51 st. Wartość wskaźnika IS = 3.61. Poniżej wykresów zamieszczono wartości współczynników predykcji sygnału oryginalnego a_0 i zniekształconego a_C.

Poniżej zaprezentowano, wyznaczone z wykorzystaniem aplikacji symulacyjnej, mapy zmienności wskaźników odległości (IS, CD, melCD) w funkcji położenia punktu odsłuchu dla trzech typów rozkładów źródeł: układ szyku źródeł w jednej linii (Dodatek B1), układ ciągu komunikacyjnego (Dodatek B2), układ matrycowy - sala audytoryjna (Dodatek B3). Symulacje przeprowadzono dla sześciu samogłosek polskich w różnych układach i rozstawach źródeł oraz w różnych płaszczyznach odsłuchu.

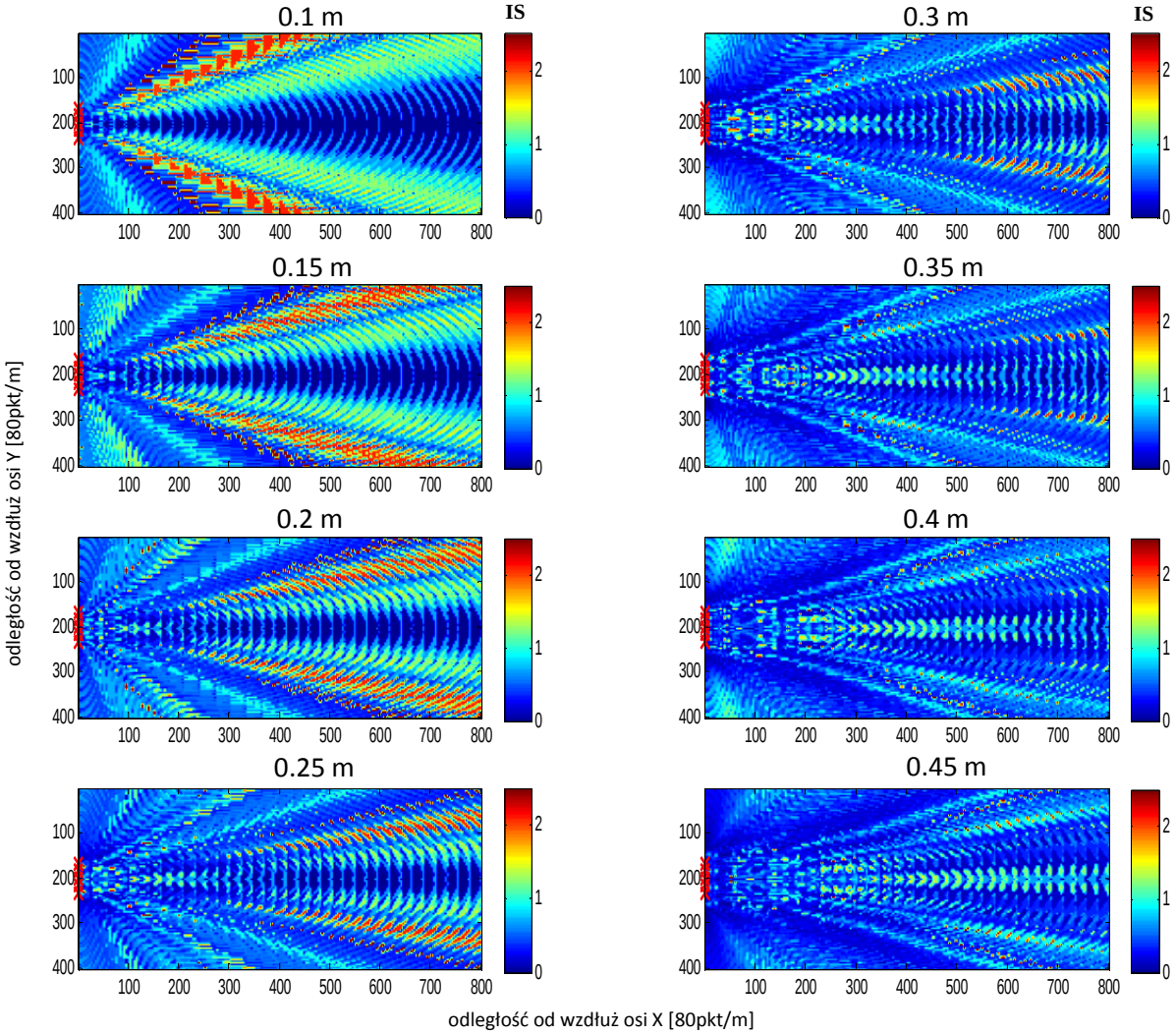
DODATEK B1

Układ akustyczny typu szyk źródeł w jednej linii.

Poniższe mapy zmienności wskaźników odnoszą się do obszaru o wymiarach 5m x 10m. Płaszczyzna odsłuchu znajduje się poniżej płaszczyzny źródeł. Zmiana rozstawu źródeł w zakresie od 0.1m do 0.45m.



Rysunek B1.1 Mapy zmienności wskaźnika melCD dla sześciu samogłosek, dla liniowego układu 5 źródeł typu szyk źródeł w jednej linii, rozstaw źródeł 0.2 m (źródła idealne).

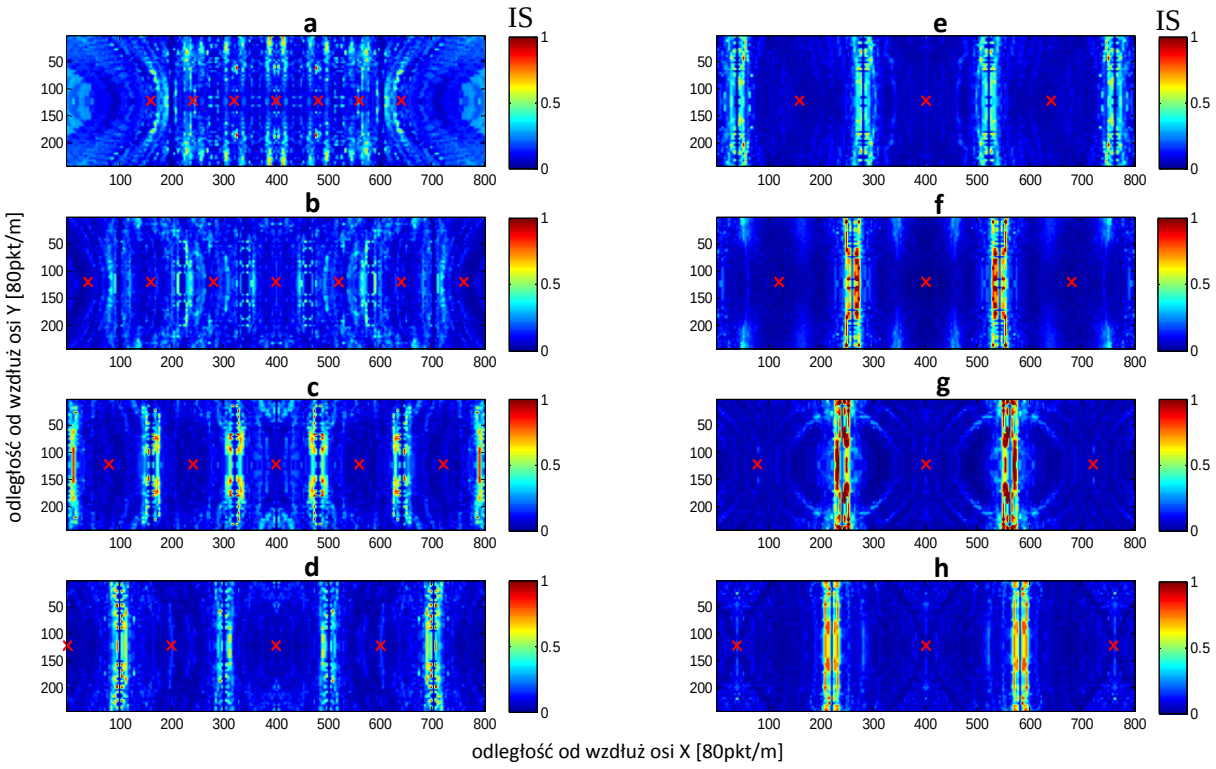


Rysunek B1.2 Mapy zmienności wskaźnika IS dla głoski „e”, dla liniowego układu 5 źródeł typu szyk źródeł w jednej linii, dla ośmiu rozstawów źródeł w zakresie od 0.2 m do 0.45 m (źródła idealne).

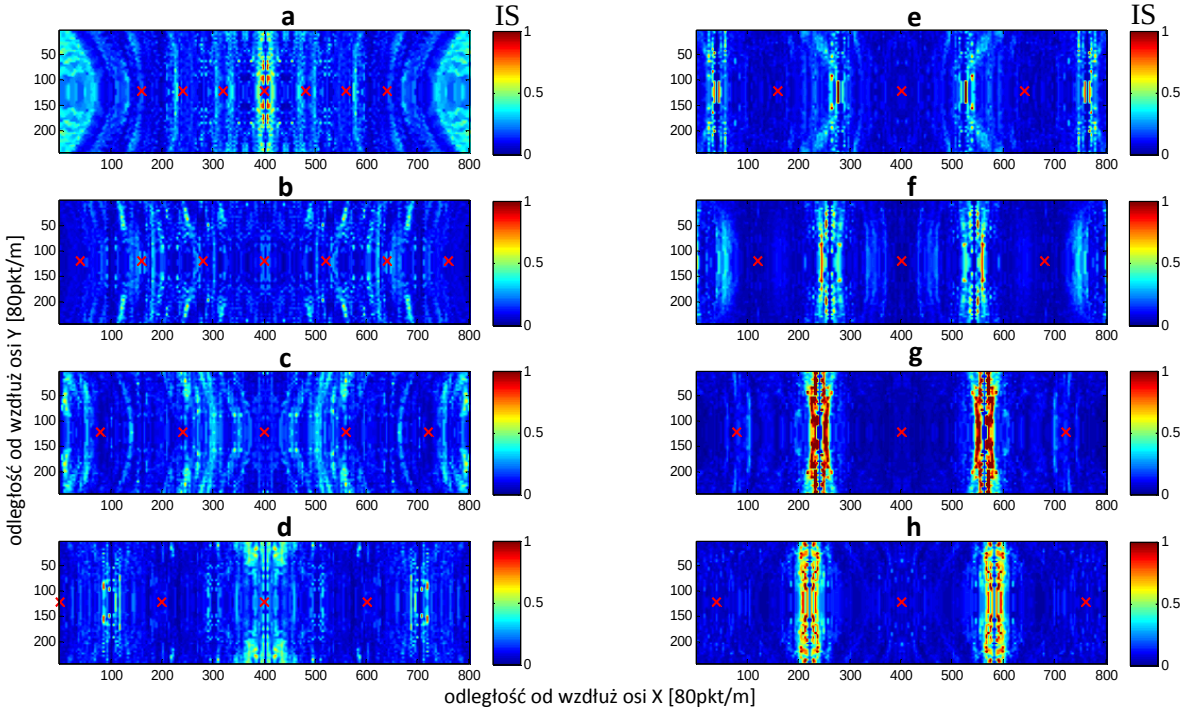
DODATEK B2

Układ akustyczny typu ciąg komunikacyjny.

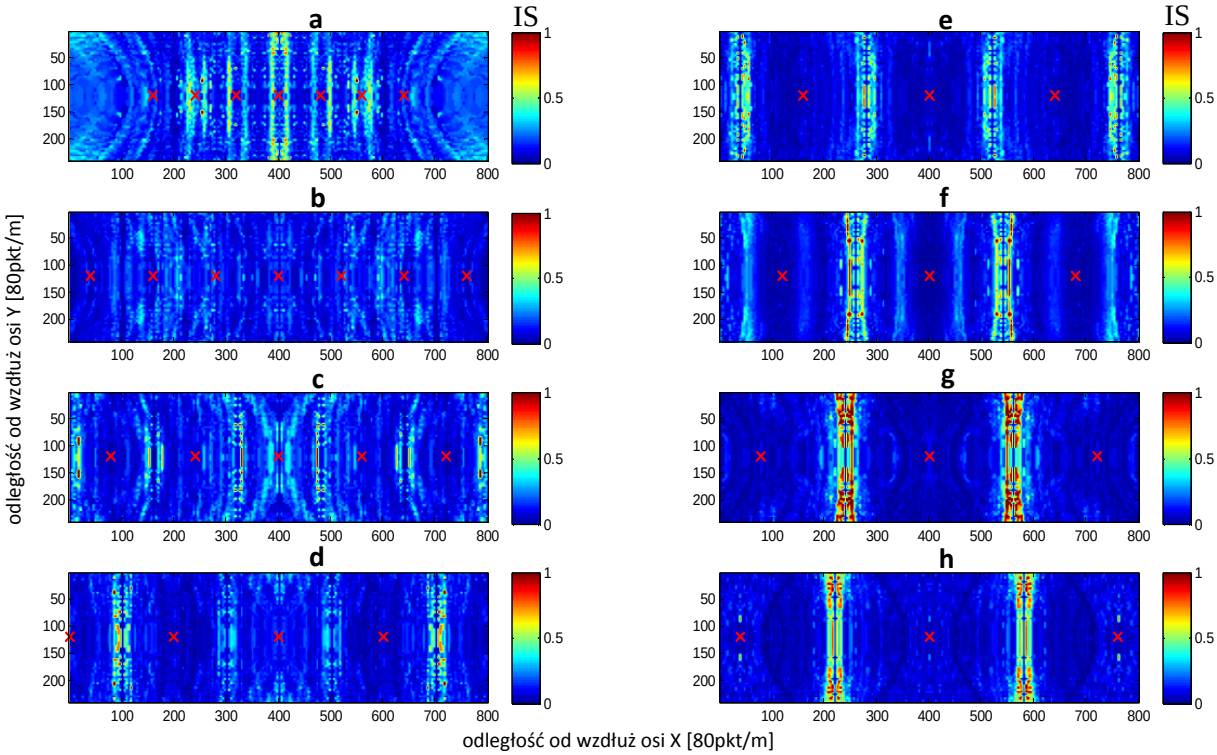
Poniższe mapy zmienności wskaźników odnoszą się do obszaru o wymiarach 3m x 10m. Płaszczyzna odsłuchu znajduje się poniżej płaszczyzny źródeł. Zmiana rozstawu źródeł 1m - 4.5m.



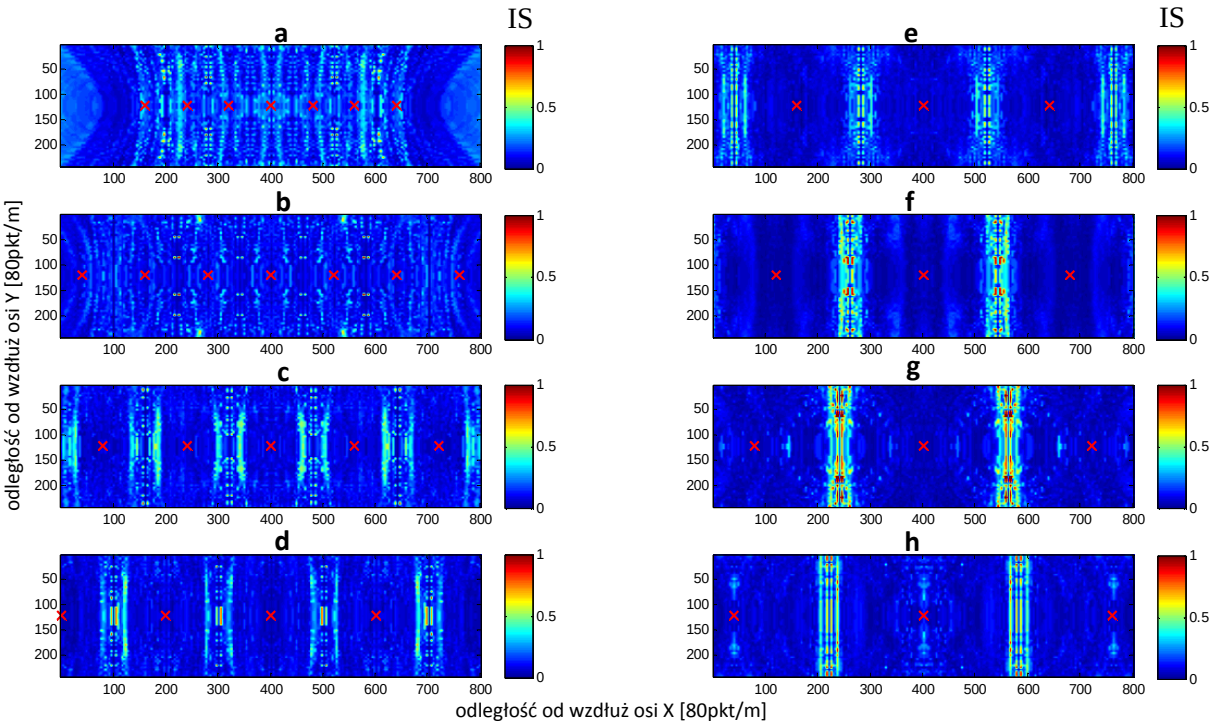
Rysunek B2.1 Mapy zmienności wskaźnika IS dla głoski „a”, dla układu typu ciąg komunikacyjny (źródła idealne). Płaszczyzna odsłuchu znajduje się 1m poniżej płaszczyzny źródeł. Zmiana rozstawu źródeł od 1m (a) do 4.5 m (h).



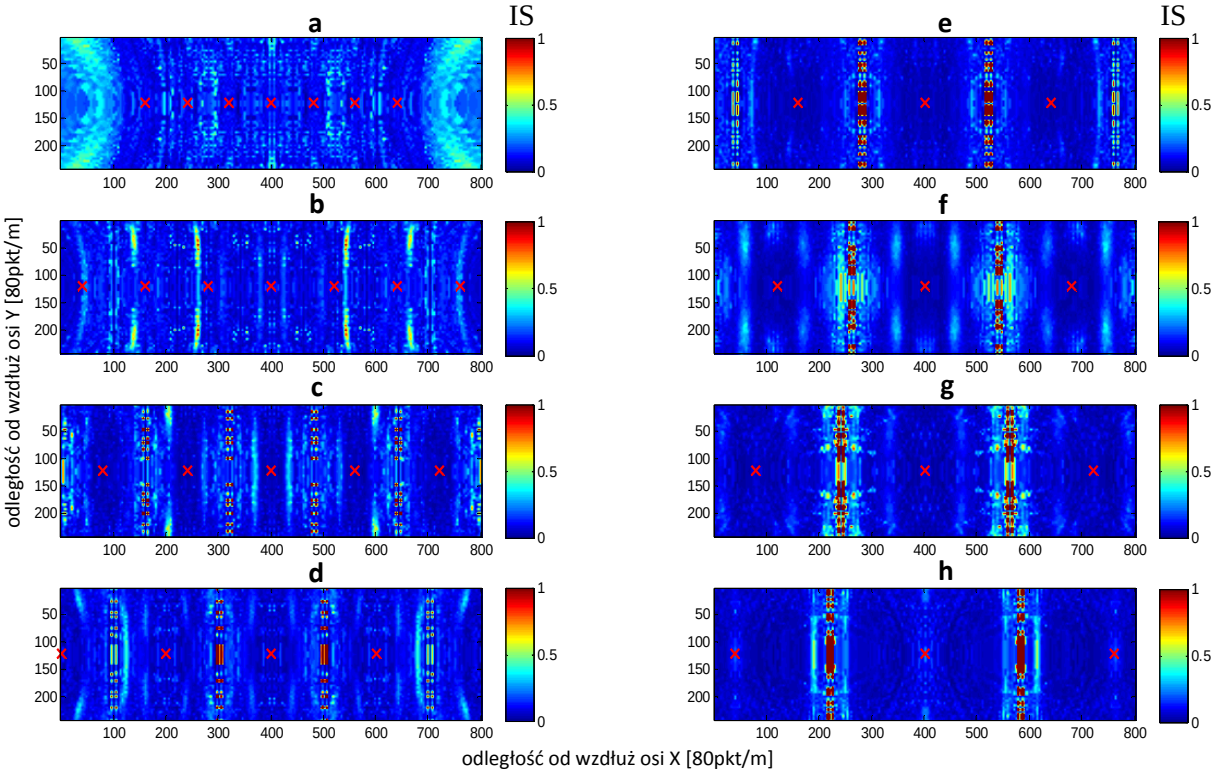
Rysunek B2.2 Mapy zmienności wskaźnika IS dla głoski „a”, dla ciąg komunikacyjny o rozmiarach 3x10 m (źródła idealne). Płaszczyzna odsłuchu znajduje się 2m poniżej płaszczyzny źródeł. Zmiana rozstawu źródeł w zakresie od 1m (a) do 4.5 m (h).



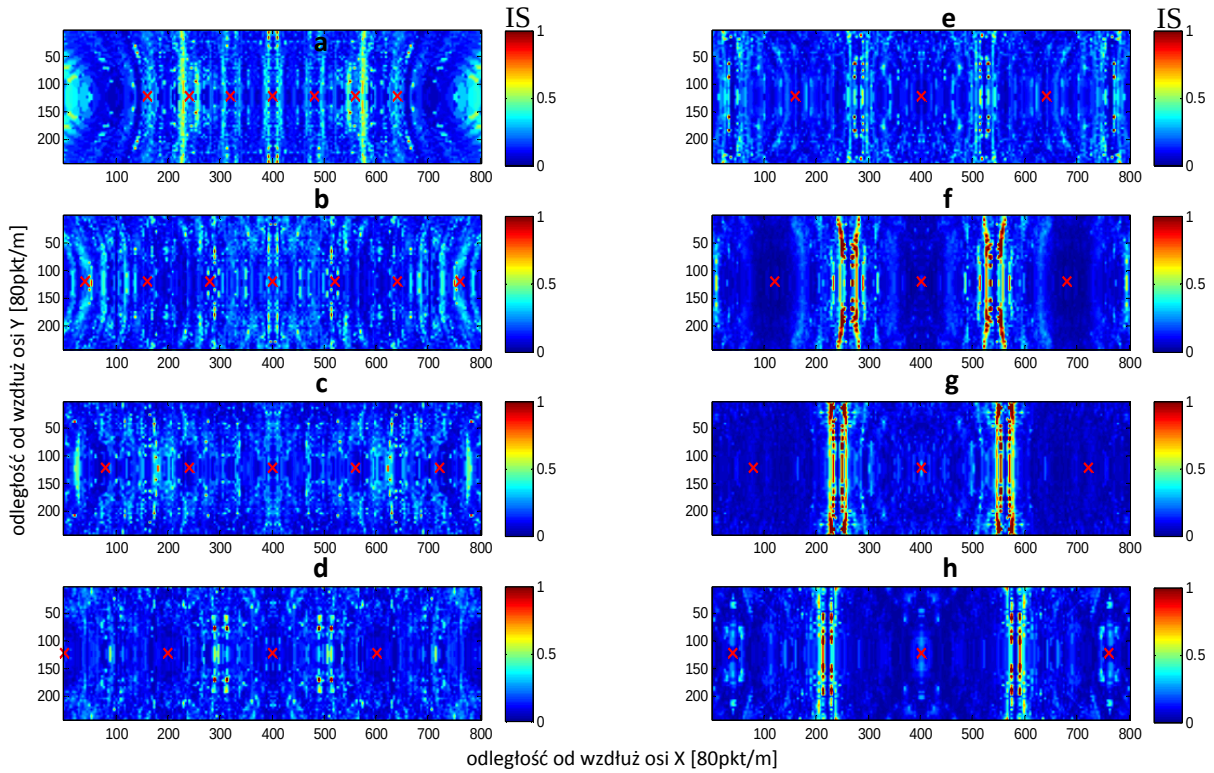
Rysunek B2.3 Mapy zmienności wskaźnika IS dla głoski „a”, dla obszaru symulującego ciąg komunikacyjny o rozmiarach 3x10 m (źródła idealne). Płaszczyzna odsłuchu znajduje się 1.5m poniżej płaszczyzny źródeł. Zmiana rozstawu źródeł w zakresie od 1m (a) do 4.5 m (h).



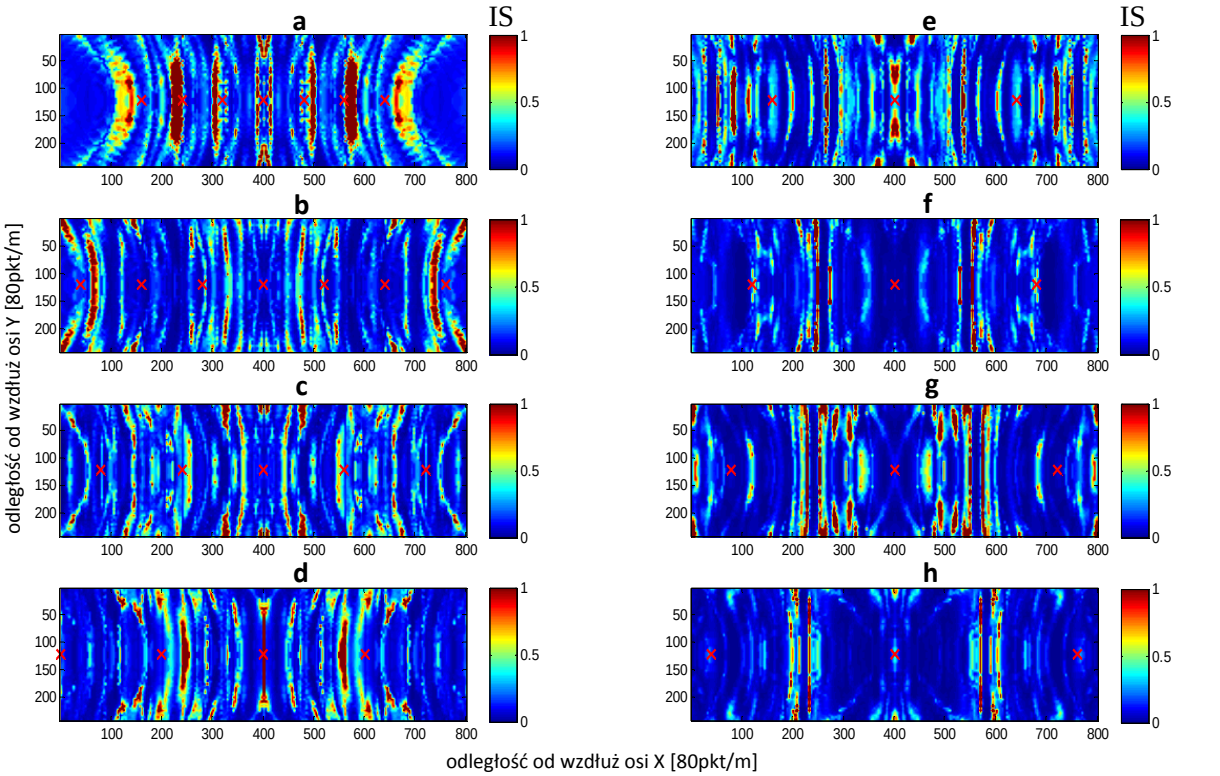
Rysunek B2.4 Mapy zmienności wskaźnika IS dla głoski „e", dla układ typu ciąg komunikacyjny – źródła idealne. Płaszczyzna odsłuchu znajduje się 1.5m poniżej płaszczyzny źródeł. Zmiana rozstawu źródeł w zakresie od 1m (a) do 4.5 m (h).



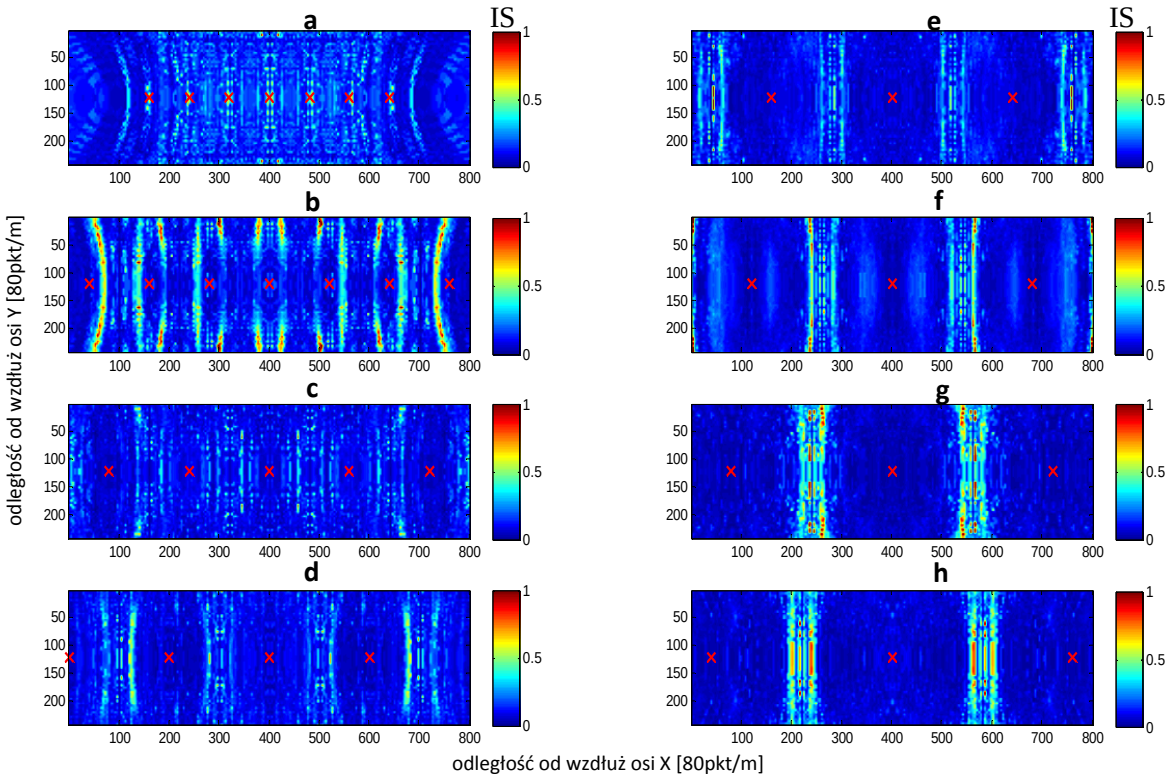
Rysunek B2.5 Mapy zmienności wskaźnika IS dla głoski „i", dla układ typu ciąg komunikacyjny – źródła idealne. Płaszczyzna odsłuchu znajduje się 1.5m poniżej płaszczyzny źródeł. Zmiana rozstawu źródeł w zakresie od 1m (a) do 4.5 m (h).



Rysunek B2.6 Mapy zmienności wskaźnika IS dla głoski „o”, dla układ typu ciąg komunikacyjny – źródła idealne. Płaszczyzna odsłuchu znajduje się 1.5m poniżej płaszczyzny źródeł. Zmiana rozstawu źródeł w zakresie od 1m (a) do 4.5 m (h).



Rysunek B2.7 Mapy zmienności wskaźnika IS dla głoski „u”, dla układ typu ciąg komunikacyjny – źródła idealne. Płaszczyzna odsłuchu znajduje się 1.5m poniżej płaszczyzny źródeł. Zmiana rozstawu źródeł w zakresie od 1m (a) do 4.5 m (h).

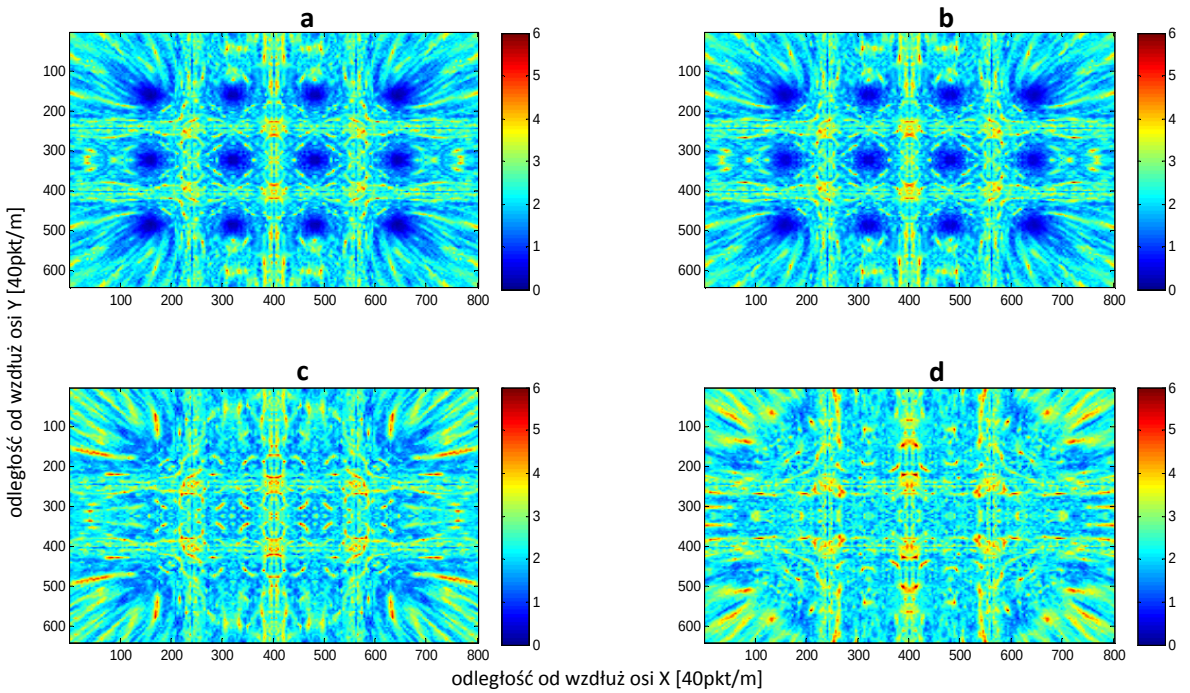


Rysunek B2.8 Mapy zmienności wskaźnika IS dla głoski „y”, dla układ typu ciąg komunikacyjny – źródła idealne. Płaszczyzna odsłuchu znajduje się 1.5m poniżej płaszczyzny źródeł. Zmiana rozstawu źródeł w zakresie od 1m (a) do 4.5 m (h).

DODATEK B3

Układ akustyczny typu sala audytoryjna

Poniższe mapy zmienności wskaźników odnoszą się do obszaru o wymiarach podłogi 16m x 20m, nagłaśnianego przez 12 źródeł rozmieszczonych w jednej płaszczyźnie (np. w suficie). Źródła rozmieszczone są równomiernie co 4m wzdłuż obu osi. Środek apertury promieniującej znajduje się nad środkiem płaszczyzny odsłuchu. Płaszczyzna odsłuchu znajduje się poniżej płaszczyzny źródeł.



Rysunek B3.1 Mapy zmienności wskaźnika odległości CD (Cepstrum Distance) w funkcji położenia punktu odsłuchu. Obszar o wymiarach podłogi 16m x 20m nagłaśniany jest przez 12 źródeł rozmieszczonych w jednej płaszczyźnie (np. w suficie). Źródła rozmieszczone są równomiernie co 4 m wzdłuż obu osi. Środek apertury promieniującej znajduje się nad środkiem płaszczyzny odsłuchu. Płaszczyzna odsłuchu znajduje się poniżej płaszczyzny źródeł o: a) 0.1 m; b) 0.2 m; c) 1 m; d) 2 m. W celach porównawczych wartości wskaźnika CD dla wszystkich przypadków zostały ograniczone do jednakowego zakresu [0, 6].