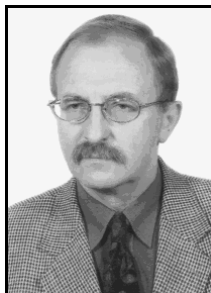


Henryk LASOTA, Romuald MAZUREK

POLITECHNIKA GDAŃSKA, WYDZIAŁ ELEKTRONIKI, TELEKOMUNIKACJI I INFORMATYKI

Modelowanie i pomiary nagłośnienia audytorium**Doc. dr hab. inż. Henryk LASOTA**

Absolwent Wydziału Elektroniki Politechniki Gdańskiej (1969), specjalista w dziedzinie akustyki szerokopasmowej, anten hydroakustycznych oraz wysokorozdzielczych systemów umożliwiających identyfikację obiektów podwodnych. Autor i współautor publikacji z zakresu analizy problemów dyfrakcji fal szerokopasmowych metodą przestrzennie - czasowych odpowiedzi impulsowych.

e-mail: henryk.lasota@eti.pg.gda.pl**Mgr inż. Romuald MAZUREK**

Absolwent Wydziału Elektroniki, Telekomunikacji i Informatyki Politechniki Gdańskiej (2001). Przygotowuje rozprawę doktorską z zakresu telekomunikacji, na temat wpływu przestrzennie-czasowych rozkładów pola na jakość komunikacji w szerokopasmowych systemach akustycznych.

e-mail: romuald.mazurek@eti.pg.gda.pl**Streszczenie**

Uwaga projektantów systemów nagłośnieniowych zogniskowana jest na tzw. akustyce pomieszczenia, w którym realizowany jest przekaz słowny. Zjawiska pogłosowe związane z geometrią sali audytorialnej, wynikające z wielokrotnych odbić fali dźwiękowej uznawane są za dominujące. Artykuł podejmuje niedoceniany problem wpływu liczby i rozmieszczenia głośników w sali na jakość przekazu. Superpozycji sygnałów dochodzących do słuchacza z wielu źródeł towarzyszy zjawisko interferencji, które prowadzi 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 funkcjami przenoszenia. Zaprezentowano wyniki symulacji i pomiarów, zwracając uwagę na wpływ, jaki zwielokrotnianie źródeł dźwięku ma na formanty - parametry fonemów istotne dla zrozumiałości mowy.

Słowa kluczowe: systemy nagłośnieniowe, kanał komunikacyjny, odpowiedź impulsowa kanału, metoda MLS, zrozumiałość mowy.

Modelling and measurements of public address system**Abstract**

Public address systems are designed with attention focused on acoustic features of the auditory hall. Reverberation phenomena related to the hall geometry, coming from multiple reflections of the sound wave are taken for dominating ones. The paper undertakes the underestimated problem of the influence of the number and distribution of loudspeakers on the communication quality. Linear superposition of signals reaching the listener from numerous sources is accompanied by interference effects that lead to deep modification of received signals. Distortions introduced by public address system has been analyzed with the use of linear systems model in which transmission features of communication systems are characterized by impulse response functions and transfer functions. Some results of simulation and measurements has been presented, showing the influence of the use of multiple sound sources on formants - phoneme parameters important for speech intelligibility.

Keywords: public address systems, communication channel, channel impulse response, MLS technique, speech intelligibility.

1. Wprowadzenie

Sluchanie mowy dostatecznie głośnej w stosunku do zakłóceń (szmery, hałasy) panujących w miejscu odsłuchu, pozwala rozumieć ją bez specjalnego wysiłku. Systemy nagłaśniania instalowane w miejscach publicznych, określane w literaturze anglojęzycznej mianem Public Address Systems, dostosowują poziom głośności do potrzeb przez instalowanie dodatkowych głośników w pobliżu słuchaczy. Mimo to w licznych przypadkach daje się zauważyć, że zrozumienie wzmocnionego przekazu nadal wymaga wyjątkowej uwagi – co prawda jest wystarczająco głośno, lecz w zamian nie dość wyraźnie.

Artykuł prezentuje podejście do problemu wykorzystujące formalizm analizy systemów liniowych. Autorzy traktują wielogłośnikowy system nagłaśniający jako szczególnego rodzaju szerokopasmowy system komunikacyjny, w którym wielodrożna transmisja prowadzi do specyficznych zniekształceń interferencyjnych sygnału mowy, mogących w wielu obszarach odsłuchu znacząco utrudnić zrozumienie przekazu. Zjawiska tego nie można pomijać przy projektowaniu systemów nagłaśniających.

Postawiony w artykule problem jest komplementarny do klasycznych zagadnień akustyki architektonicznej, która kłopoty z „akustyką” w salach traktuje li tylko w kategoriach pogłosu, jako wynikające z geometrii pomieszczeń i występujących w nich odbić fali dźwiękowej.

2. Nagłośnienie jako system komunikacyjny

Najprostszy system nagłośnienia składa się z mikrofonu, układów wzmacniających i pojedynczego głośnika będącego źródłem fali dźwiękowej. Typowe systemy publicznego przekazu słownego 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). Fala dźwiękowa docierająca do dowolnego punktu w przestrzeni odsłuchowej jest wówczas liniową superpozycją fal promieniowanych przez poszczególne źródła.

Chociaż zagadnienie formowania się pola akustycznego 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 z powodzeniem traktować jako czysto harmoniczne, natomiast sygnał mowy jest z natury szerokopasmowy, jego przebieg czasowy jest złożony a widmo częstotliwościowe pokrywa 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.

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. pole dalekie).

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, należy postrzegać jako realizację 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.

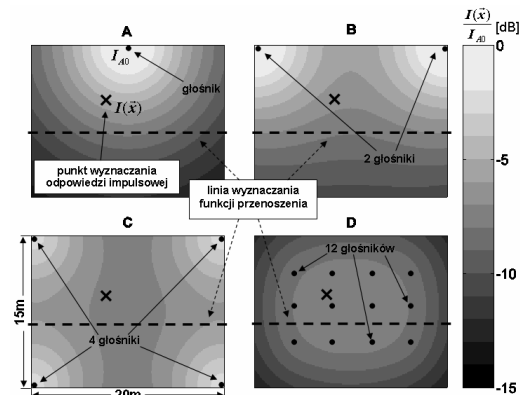
Problemy z bezbłędnym odbiorem komunikatu są w równym stopniu rezultatem wpływu czynników obiektywnych, które można uchwycić badając zniekształcające właściwości kanału transmisyjnego oraz wpływ zakłóceń, co trudnych do jednoznacznego scharakteryzowania czynników subiektywnych, wynikających ze złożoności 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 umożliwi 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.

3. Model optyczny

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

Dla zilustrowania problemów związanych z formowaniem pola przez systemy nagłaśniające, obliczenia przedstawione w artykule wykonano dla hipotetycznego audytorium o wymiarach 15 m x 20 m, z doskonale pochłaniającymi powierzchniami ograniczającymi, bez wyposażenia (rys.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 nagłaśnione 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, rozkład natężeń okazał się najbardziej równomierny dla największej liczby źródeł rozmieszczonych nad płaszczyzną odsłuchu.



Rys. 1. Rozkład natężenia dźwięku w modelowym audytorium nagłaśnionym nieskorelowanymi falami generowanymi przez: A) pojedyncze źródło, B) dwa źródła, C) cztery źródła, D) 12 źródeł

Fig. 1. Sound intensity distribution in model auditory hall insonified with uncorrelated waves generated by: A) single source, B) two sources, C) four sources, D) 12 sources

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 nagłaśnieniowego.

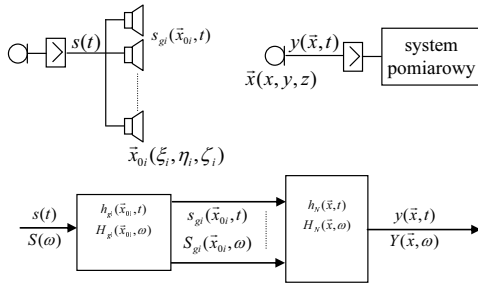
4. Model liniowy

Podstawowym fizycznym mechanizmem formowania pola akustycznego w przestrzeni jest zachodząca w każdym jej punkcie liniowa superpozycja chwilowych wartości ciśnienia. Gdy pole dźwiękowe jest kreowane przez źródła pobudzone niemal identycznym sygnałem, o niewielkich różnicach wynikających z indywidualnych cech głośników, dominującym zjawiskiem jest interferencja, której skutki dla sygnałów szerokopasmowych są praktycznie mało znane. Jest ona szczególnie silna w obszarach „obsługiwanych” 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 audytorium i, w konsekwencji, pogorszenie zrozumiałości przekazu.

Dla analizy problemu interferencji sygnałów szerokopasmowych zakładamy model systemowo – liniowy, w którym 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). Dla uproszczenia pomijamy kierunkowość głośników, przyjmując, że mają one własność źródeł punktowych.

W każdym punkcie nagłaśnianej przestrzeni, N -głośnikowy system można scharakteryzować 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 transformata Fouriera odpowiedzi impulsowej, charakteryzuje jego własności bezpośrednio w stanach ustalonych.





Rys. 2. System nagłośnienia (a) i jego model systemowo-liniowy (b)
Fig. 2. Public address system (a) and its linear system model (b)

Właściwości transmisyjne swobodnej przestrzeni trójwymiarowej opisać można funkcją odpowiedzi impulsowej $h(r,t)$ o symetrii sferycznej. Charakteryzuje ona system liniowy złożony ze źródła punkowego i z punktu odsłuchu znajdującego się w odległości r od źródła, w sposób następujący:

$$h(r,t) = \frac{r_1}{r} \delta(t - r/c), \tag{1}$$

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

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

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:

$$h_N(\vec{x},t) = \sum_{i=1}^N h_i(r_i,t), \tag{3}$$

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

$$r_i = \left[(x - \xi_i)^2 + (y - \eta_i)^2 + (z - \zeta_i)^2 \right]^{1/2}. \tag{4}$$

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 spłotową w dziedzinie czasu (oznaczoną symbolem $*$) sygnału z odpowiedzią głośnika:

$$s_{gi}(t) = s(t) * h_{gi}(t). \tag{5}$$

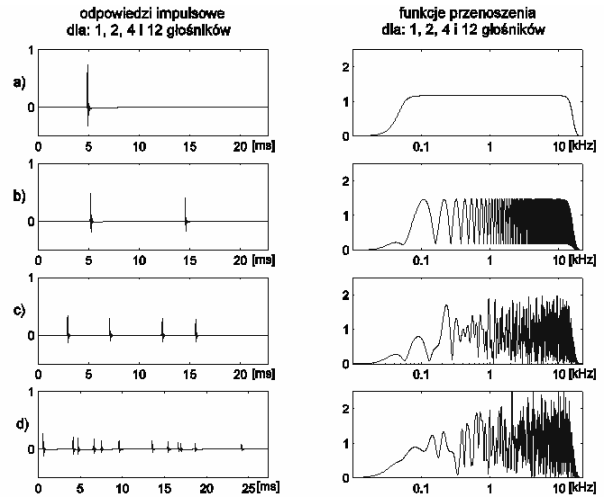
Odpowiedź impulsowa łańcucha komunikacyjnego złożonego z N różnych głośników ma więc formę:

$$h_{Ng}(\vec{x},t) = \sum_{i=1}^N [h_{gi}(t) * h_i(r_i,t)]. \tag{6}$$

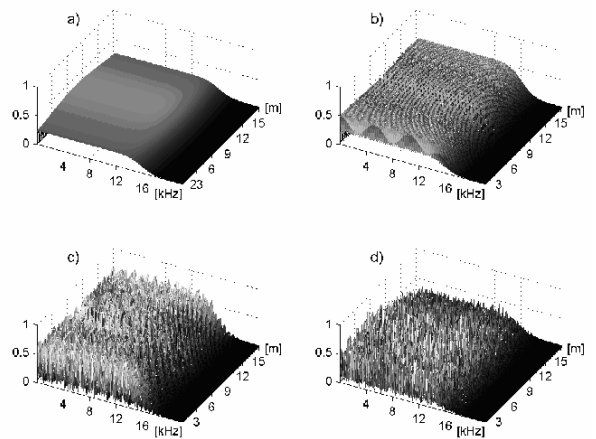
Jeśli założyć, że głośniki są jednakowe, o odpowiedziach $h_{gi}(t) = h_g(t)$, odpowiedź systemu można zapisać w postaci:

$$h_{Ng}(\vec{x},t) = h_g(t) * \sum_{i=1}^N h_i(r_i,t) = h_g(t) * h_N(\vec{x},t). \tag{7}$$

Rys. 3 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.1 - z jednym, dwoma, czterema i 12 głośnikami, w punkcie odsłuchu X. Rys. 4 przedstawia moduł funkcji przenoszenia obliczony dla kolejnych punktów leżących wzdłuż jednej linii płaszczyzny odsłuchu, w stałej odległości od ściany frontowej dla czterech konfiguracji. W obu przypadkach założono, że głośniki są źródłami punktowymi o płaskiej charakterystyce częstotliwościowej w zakresie od 50 Hz do 15 kHz (zobrazowanej na rys. 3 a wraz z odpowiedzią impulsową modelowego głośnika).



Rys. 3. Odpowiedź impulsowa i funkcja przenoszenia symulowanego systemu nagłośnienia w przykładowym punkcie odsłuchu jak na rys.1: a) 1 źródło, b) 2 źródła, c) 4 źródła, d) 12 źródeł; logarytmiczna skala częstotliwości.
Fig. 3. Impulse response and transfer function of simulated public address system in the audition point as in Fig. 1: a) 1 source, b) 2 sources, c) 4 sources, d) 12 sources; logarithmic frequency scale



Rys. 4. 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
Fig. 4. Transfer function modulus of simulated public address system in the audition points along $x=8.7$ m line: a) 1 source, b) 2 sources, c) 4 sources, d) 12 sources; linear frequency scale

Otrzymane wyniki obliczeniowe pozwalają ocenić degradujący wpływ zjawiska interferencji szerokopasmowej na właściwości transmisyjne. Warto zauważyć, że kanał akustyczny nie modyfikuje właściwości 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:

$$y(\vec{x}, t) = s(t) * h_{Ng}(\vec{x}, t). \quad (8)$$

Przedmiotem zainteresowania przy ocenie jakości systemu jest deformacja odbieranego sygnału w stosunku do sygnału pobudzającego, zmniejszająca prawdopodobieństwo prawidłowej detekcji fonemów.

5. Pomiary odpowiedzi systemu

Występowanie efektów zaobserwowanych w obliczeniach symulacyjnych zweryfikowano wykonując pomiary odpowiedzi impulsowych w audytorium Wydziału Elektroniki, Telekomunikacji i Informatyki Politechniki Gdańskiej, o rozmiarach podobnych do przyjętych w modelu z rys. 1. W pomiarach wykorzystano metodę korelacyjną z zastosowaniem sygnałów MLS, zalecaną przez standardy ISO dla akustyki pomieszczeń [1], mającą istotną przewagę nad metodą bezpośrednią.

Idea pomiaru korelacyjnego odpowiedzi impulsowej zasadza się na istotnym związku, jaki w systemach liniowych istnieje między funkcją autokorelacji $\rho_{ss}(t)$ sygnału pobudzającego $s(t)$ i funkcją kroskorelacji $\rho_{sy}(t)$ sygnałów pobudzającego i wyjściowego $y(t)$.

Można mianowicie wykazać, że:

$$\rho_{sy}(t) = h(t) * \rho_{ss}(t), \quad (9)$$

co oznacza, że jeśli system pobudzić sygnałem o wąskiej, podobnej do dystrybucji delta Diraca funkcji autokorelacji, to korelacja skrośna sygnałów wyjściowego i wejściowego jest dobrym przybliżeniem odpowiedzi impulsowej badanego systemu:

$$\rho_{sy}(t) \approx h(t). \quad (10)$$

W metodach korelacyjnych najlepiej spisują się generowane cyfrowo binarne sygnały bipolarnie, tzw. MLS generowane na bazie pseudo-przypadkowych tzw. m-ciągów znanych jako ciągi maksymalnej długości (Maximum Length Sequences) oraz sygnały zwane świergotowymi, czyli sinusoidy z liniowym przestrajaniem częstotliwości (*ang.* chirp).

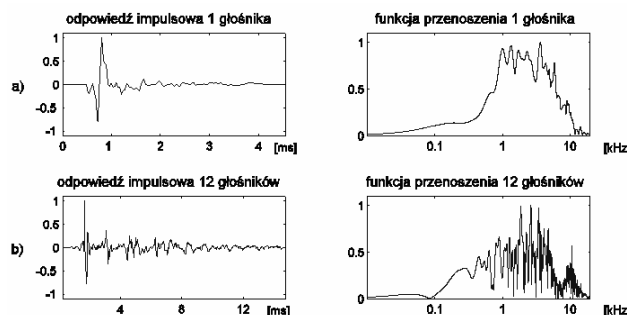
Metoda MLS daje wyniki z wysokim stosunkiem sygnału do zakłóceń, pod warunkiem zachowania pełnej synchronizacji układu odbiorczego z układem nadawczym. Zaleca się więc, by w pomiarach używać tego samego komputera, zarówno do obróbki *in situ*, jak do obliczeń wykonywanych po pomiarach.

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 mowy, do wzmacniaczy zasilających zespół 12 głośników wmontowanych w sufit. 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 - Fast Hadamard Transform) [2].

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. 2) oraz N-krotnego kanału akustycznego (przestrzeń między kolejnymi głośnikami i punktem pomiaru). Przetworzenie odpowiedzi w dziedzinie częstotliwości przy pomocy przekształcenia Fouriera daje funkcję przenoszenia systemu.

Rysunek 5 przedstawia przykładowe wyniki pomiarów przeprowadzonych w audytorium w odniesieniu do pojedynczego głośnika (pomiar wykonany z odległości 1 m) oraz całości syste-

mu. W pierwszym przypadku można uznać, że zarejestrowano „czystą” odpowiedź impulsową głośnika, przed dotarciem do punktu pomiarowego sygnałów z pozostałych głośników oraz sygnałów pogłosowych (odbicie od powierzchni ograniczających i od wyposażenia - pulpitów, siedzeń). W drugim przypadku mieliśmy do czynienia z pełną odpowiedzią systemu (sygnały z 12 głośników). Do analizy wybrano początkowy fragment odpowiedzi impulsowych, tym samym eliminując wpływ pogłosu.



Rys. 5. Odpowiedź impulsowa oraz charakterystyka częstotliwościowa (moduł funkcji przenoszenia) w skali logarytmicznej: a) jednego z 12 głośników systemu nagłośnienia sali audytorijnej oraz b) całego systemu
Fig. 5. Impulse response and frequency characteristics (transfer function) in logarithmic scale of: a) one of 12 loudspeakers of public address system in the auditorium and b) complete system

6. Zrozumiałość zniekształconego przekazu mowy

Problem oceny jakości przekazu mowy występuje we wszystkich systemach komunikacji głosowej. Na jakość składają się liczne parametry, takie jak głośność, naturalność, zrozumiałość. W systemach nagłośnienia kluczowa jest zrozumiałość jako decydująca o wartości informacyjnej przekazu.

Zniekształcenia przekazu powstają zarówno w komunikacji bezpośrednio, w pomieszczeniach i obiektach o tzw. „słabej akustyce” (echo, pogłos, tłumienie), jak i w systemach stosujących przeróżne środki techniczne. Dla ich badania opracowano wiele metod i standardów zalecanych do stosowania w różnych obszarach techniki.

Istnieją dwie 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 - Mean Opinion Score). Druga kategoria, to metody obiektywne, w których ocena dokonywana jest na podstawie pomiarów parametrów systemu niezależnych od słuchacza.

Ze względu na specyfikę komunikacji międzyludzkiej, większe zaufanie okazywane jest dla metod „subiektywnych”, w których rejestrowane są błędy przekazu w badanym systemie. Cenne w nich jest to, że badania odnoszą się bezpośrednio do istoty komunikacji a wyniki odzwierciedlają poziom satysfakcji adresatów przekazu. Jednak metody te są czasochłonne, kosztowne i kłopotliwe organizacyjnie, a ponadto uzyskiwane wyniki nie są w pełni powtarzalne, jako że mocno zależą od cech osobniczych zarówno mówcy jak i słuchacza - od właściwości artykulacyjnych pierwszego oraz percepcji słuchowej drugiego.

Z powyższych powodów, niegasnące zainteresowanie budzą 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. Trudność polega na tym, że związek między parametrami systemów, wprowadzanymi przez zniekształceniami sygnału mowy i wynikającą z nich degradacją zrozumiałości nie jest dostatecznie jednoznaczny. Pomimo sporego wysiłku poświęconego rozwojowi obiektywnych metod pomiaru jakości

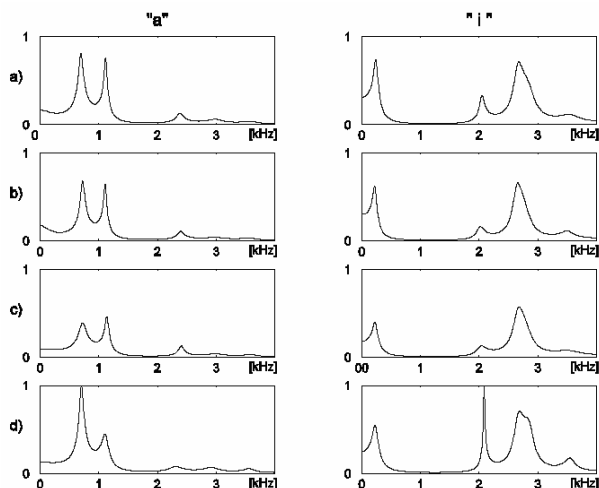
transmisji mowy, uzyskiwane wyniki nie przez wszystkich są uznawane za wystarczająco wiarygodne. Zalecane przez normy obliczenia indeksów zrozumiałości STI (Speech Transmission Index), bazujące na pomiarach tzw. funkcji przenoszenia modulacji MTF (Modulation Transfer Function), zaprojektowane dla oceny wpływu zjawisk pogłosowych na jakość przekazu, nie cieszą się pełnym zaufaniem, gdyż związek funkcji MTF ze zrozumiałością jest w istocie pośredni i niejasny [3, 4].

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. Obiektywna ocena wpływu zniekształceń liniowych na degradację zrozumiałości mowy musi odbywać się na poziomie fonemów względnie logatomów, będących elementarnymi sekwencjami zawierającymi także stany przejściowe między fonemami (tzw. alofony). Warto przy tym mieć na uwadze, że przy odsłuchu w warunkach szumowych i/lub pogłosowych, a także dla przekazu dokonywanego w języku obcym dla słuchacza, w którym fonemy mają inne brzmienie, efekt obniżenia zrozumiałości fonemowej wyodróżnia się.

Jak wiadomo, o zrozumiałości przekazu słownego nie decyduje w sposób jawny i bezpośredni ani kształt sygnału ani jego zawartość widmowa. Zrozumiałość związana jest z mechanizmem generacji głosu. Dla każdego fonemu, stan traktu głosowy mówcy daje się opisać zbiorem parametrów, które znajdują odzwierciedlenie w występowaniu tzw. formantów, czyli charakterystycznych maksimum widma chwilowego sygnału. Wzajemne położenie oraz poziom formantów występujących w fonemach decyduje o możliwości rozróżnienia tych ostatnich.

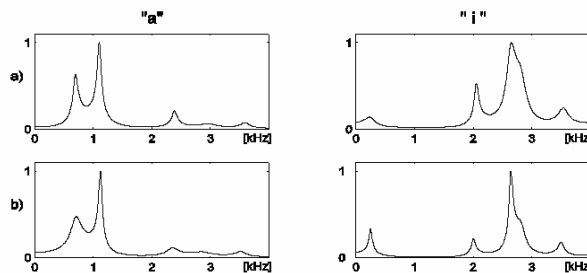
Autorzy zastosowali analizę predykcyjną w standardzie LPC-10 (Linear Predictive Coding rzędu 10), która umożliwiła ekstrakcję do 5 formantów [5]. Dla oceny zniekształceń, obliczono obwiednię widma sygnału $y(t)$ (tzw. pseudowidmo LPC) dla kilku różnych fonemów mowy polskiej, po poddaniu ich splotowi z odpowiedziami impulsowymi $h_{N_g}(t)$ systemu nagłośnienia, zarówno obliczonymi dla konfiguracji modelowych, jak i pomierzonymi w rzeczywistym audytorium.

Wyniki obliczeń i pomiarów weryfikujących wskazują, że wzrost liczby źródeł sygnału powoduje powstawanie silnych zniekształceń lokalnych maksimum obwiedni. Rys. 6, odpowiadający sytuacji z rys. 3, przedstawia pseudowidmo LPC głosek „a” oraz „i” w wybranym punkcie odsłuchu dla różnej liczby i konfiguracji źródeł. Przy wzroście liczby źródeł, od pojedynczego, poprzez 2 i 4 źródła aż do układu 12-tu źródeł, następuje wyraźna deformacja proporcji między poszczególnymi formantami. Podobny efekt zaobserwowano dla rzeczywistych pomiarów sali audytorijnej (rys. 7, sytuacja z rys. 5).



Rys. 6. Pseudowidmo LPC-10 sygnałów głosek „a” oraz „i” po przejściu przez modelowany system – konfiguracje jak na rys. 3

Fig. 6. Pseudospectrum LPC-10 of „a” and „i” phonemes at the output of modeled system - configurations as in Fig. 3



Rys. 7. Pseudowidmo LPC-10 sygnałów głosek „a” oraz „i” po przejściu przez system mierzony jak na rys. 5: a) 1 głośnik, b) 12 głośników

Fig. 7. Pseudospectrum LPC-10 of „a” and „i” phonemes at the output of measured system, as in Fig. 5: a) 1 loudspeaker, b) 12 loudspeakers

Jak się okazuje, zniekształcenia wprowadzane przez kanał transmisyjny silnie wpływają na poziom poszczególnych formantów, w mniejszym stopniu na ich położenie.

Zaobserwowane efekty potwierdzono odsłuchowo – stwierdzono degradację zrozumiałości fonemów w stosunku do odsłuchu z jednego źródła, zarówno dla przypadku symulacji (rys. 6), jak i dla pomiarów (rys. 7). Najsilniejsze zniekształcenia obserwuje się dla konfiguracji 12 głośników, dla której analiza natężeniowa dała najlepszy jakościowo wynik (rys. 1.).

7. Podsumowanie

Zaprezentowane w artykule symulacja i pomiary systemów nagłośnieniowych wykazały, że zjawisko interferencji towarzyszące generowaniu pola dźwiękowego przez źródła pobudzone identycznym sygnałem powoduje znaczne zniekształcenia liniowe. Jakość przekazu słownego ulega pogorszeniu, zależnemu w silnym stopniu od liczby źródeł dźwięku oraz od ich rozmieszczenia względem miejsca odsłuchu.

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

Badania realizowane przez autorów mają na uwadze opracowanie narzędzi analizy dostosowanych do specyfiki problemu formowania pola przez zespół nieregularnie rozmieszczonych źródeł szerokopasmowych. Chodzi o umożliwienie, na etapach projektowania i testowania systemów nagłośnieniowych, oceny jakości transmisji sygnałów mowy z punktu widzenia zachowania odpowiedniego poziomu zrozumiałości przekazu. Analiza przedstawiona w artykule powinna doprowadzić do opracowania zaleceń uzupełniających już istniejące.

8. Literatura

- [1] ISO 3382, Acoustics - Measurement of the reverberation time of rooms with reference to other acoustical parameters, 1997.
- [2] D.D.Riffe, J.Vanderkooy: Transfer-Function Measurement with Maximum-Length Sequences, J.Audio Eng. Soc., vol. 37, pp. 419-443, 1989.
- [3] T. Houtgast, H. Steeneken: The modulation transfer function in room acoustics as a predictor of speech intelligibility. Acustica 28, p. 66, 1973.
- [4] PN-EN 60268-16, Urządzenia systemów elektroakustycznych. Cz.16: Obiektywna ocena zrozumiałości mowy za pomocą wskaźnika transmisji mowy, 2005.
- [5] T.P. Zieliński: Cyfrowe przetwarzanie sygnałów, WKŁ, Warszawa 2005.