

Eugeniusz KORNATOWSKI¹

PRZESTRZENNA IDENTYFIKACJA KIERUNKU ŹRÓDŁA DŹWIĘKU

W artykule opisano możliwości wykorzystania jednopunktowego mikrofonu surround do wyznaczania kierunku źródła dźwięku (Sound Source Localization - SSL). W eksperymentach zastosowano mikrofon „soundfield” z czterema przetwornikami (kapsułami) o charakterystykach kardioidalnych. Specjalna konstrukcja mechaniczna powoduje, że przetwornik w całości posiada charakterystykę dookólną. Mikrofon ten umożliwia akwizycję dźwięku 3-D w tzw. A – formacie. Po dalszej obróbce sygnału możliwe jest wyznaczenie np. kierunku źródła dźwięku w przestrzeni. Przeprowadzone eksperymenty dowodzą, że prosty algorytm obliczeniowy szczególnie nadaje się do realizacji w czasie rzeczywistym, a zastosowanie mikrofonu „soudfield” znacząco upraszcza konstrukcję mechaniczną systemu SSL.

SPATIAL IDENTIFICATION OF THE SOUND SOURCE DIRECTION

The paper describes options to use a single point surround microphone to determine a direction of the sound source localization (SSL). A “soundfield” microphone with four transducers (capsules) characterized by cardioid responses. A unique mechanical design of the transducer results in its omnidirectional response. The microphone enables 3D sound acquisition in so called A-format. Upon further processing of signal it is possible to determine e.g. a direction of a sound source within space. The conducted experiments prove that a simple calculation algorithm is in particular feasible for a real time operation, and application of the „soudfield” microphone significantly simplifies mechanical design of the SSL system.

1. WSTĘP

Jednym z pierwszych urządzeń do lokalizacji źródeł dźwięku był opatentowany przez A. M. Mayera w 1880 roku tzw. „Topophon”. Urządzenie to miało wspomagać pracę nawigatora przy lokalizacji statków płynących wzdłuż rzeki. Konstrukcja ta przypominała używany przez lekarzy stetoskop z tą różnicą, że w miejsce „mikrofonu” zastosowano dwie tuby rozstawione we wzajemnej odległości ok. 1 m. Podobne urządzenie skonstruowano w

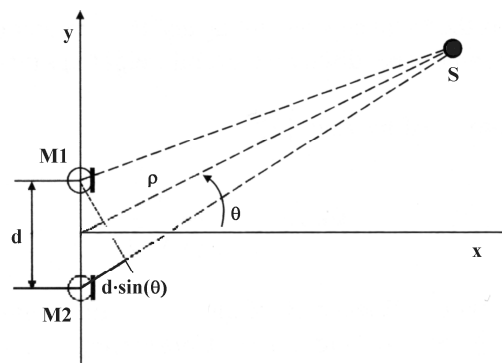
¹Zachodniopomorski Uniwersytet Technologiczny w Szczecinie, Wydział Elektryczny, ul. 26 Kwietnia 10, 71-126 Szczecin, , e-mail: korn@ps.pl.

1938 roku z przeznaczeniem do lokalizacji samolotów. Konstrukcję tę wykorzystywało brytyjskie lotnictwo.

We współczesnych konstrukcjach np. robotów (w tym mobilnych) wykorzystuje się zaawansowane techniki rozpoznawania otaczającego środowiska i obiektów. Techniki te opierają się na zastosowaniu algorytmów cyfrowego przetwarzania i analizy obrazu, a także, coraz częściej, wykorzystuje się do tych celów analizę sygnału akustycznego² [1, 2]. Szczególnie interesujące w tym zakresie są prace [3] i [4]. W każdym z proponowanych rozwiązań wykorzystuje się matrycę mikrofonów o indywidualnej konstrukcji i właściwościach. Opisywane w literaturze rozwiązania charakteryzują się, wg ich autorów, dobrymi właściwościami, umożliwiając stosunkowo dokładną lokalizację obiektów emitujących dźwięk.

2. DWUWYMIAROWY SYSTEM SSL

Najprostszy system rozpoznawania horyzontalnego kierunku dźwięku składa się z dwóch mikrofonów o dookólnej charakterystyce kierunkowej, zainstalowanych na poziomym odcinku, we wzajemnej odległości d (rys. 1.)



Rys. 1. Geometria dwuwymiarowego systemu SSL

Kierunek źródła dźwięku może być estymowany zgonie z następującą zależnością:

$$\theta = \arcsin\left(\frac{\tau_D v}{d}\right) \quad (1)$$

gdzie: v - prędkość dźwięku w powietrzu (340m/s), τ_D - różnica czasów, po których fala akustyczna dociera do mikrofonu M1 i M2. Zakłada się przy tym, że odległość ρ źródła dźwięku S od mikrofonów jest znacznie większa od wzajemnej odległości mikrofonów d . A zatem, zagadnienie rozpatruje się dla propagacji w tzw. polu dalekim. Problem lokalizacji źródła dźwięku dla jednej pary mikrofonów sprowadza się do określenia czasu

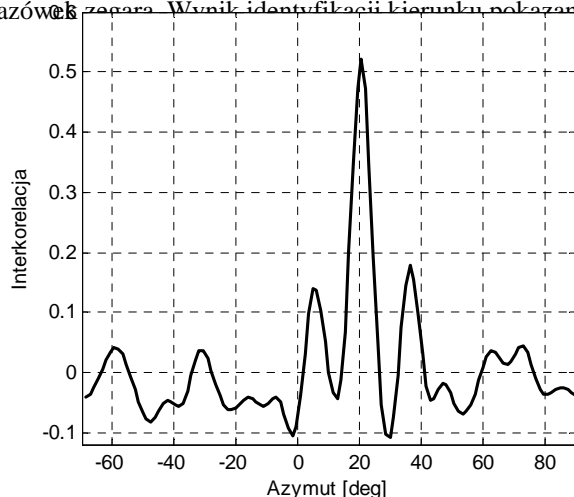
² Sound Source Localization - SSL

opóźnienia τ_D między dwoma sygnałami. Najbardziej intuicyjne, ale jednocześnie skuteczne, rozwiązanie tak postawionego problemu sprowadza się do poszukiwania maksimum funkcji interkorelacji³ dwóch dyskretnych i skończonych sygnałów $x_1(t)$ i $x_2(t)$. Jeżeli znane są transformaty Fouriera \mathbf{X}_1 i \mathbf{X}_2 tych sygnałów, to wówczas zależny od czasu wektor interkorelacji definiowany jest jako [4]:

$$\mathbf{R}_{12} = \text{iFFT}[\mathbf{X}_1 \mathbf{X}_2^H] \quad (2)$$

gdzie symbol H oznacza koniunkcję transponowaną wektora zespolonego. Zatem czas, dla którego w wektorze \mathbf{R}_{12} występuje wartość maksymalna jest poszukiwanym opóźnieniem τ_D . Dalej, wykorzystując wzór (1), z łatwością można wyznaczyć poszukiwany kierunek źródła dźwięku (kąt θ).

Stosując tę prostą koncepcję działania systemu SSL wykonano pomiary kąta azymutu pojedynczego źródła dźwięku. W odległości d wynoszącej 24 cm ustawiono dwa mikrofony typu Samson C03 o dookólnej charakterystyce kierunkowej. Jako źródło dźwięku wykorzystano zestaw głośnikowy KRK VXT8 emitujący sygnał (głos ludzki) z odbiornika radiowego. Odległość mikrofonów od głośnika wynosiła ok. 2 m. Źródło dźwięku ustawione było z niewielkim odchyleniem od osi x (rys. 1) w kierunku przeciwnym do ruchu wskazówek zegara. Wynik identyfikacji kierunku pokazano na rys. 2.



Rys. 2. Wynik identyfikacji kierunku źródła dźwięku

Położenie maksimum krzywej (odcięta) została wyliczona z uwzględnieniem wzorów (1) i (2) i wynosi ok. 20° . Uzyskany wynik odzwierciedla założone warunki eksperymentu.

System dwumikrofonowy pozwala na lokalizację źródła dźwięku na płaszczyźnie, a zatem możliwe jest określenie tylko kąta azymutu. Trójwymiarowy system SSL musi zawierać dodatkowy mikrofon (mikrofony), umożliwiając akwizycję sygnału niosącego informację o kącie elewacji. Ilość mikrofonów większa niż dwa znacznie komplikuje

³ Korelacja wzajemna, cross – correlation (ang.)

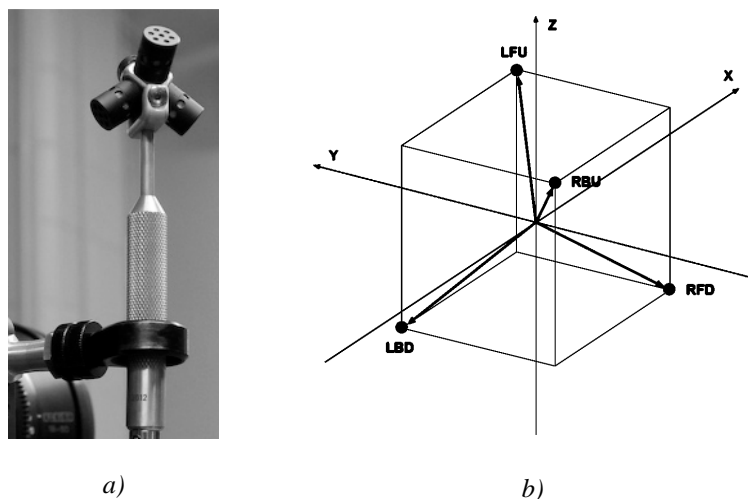
algorytm obliczeniowy, bowiem w tym przypadku należy analizować funkcję (wektor) interkorelacji między poszczególnymi parami przetworników. Ten fakt ma szczególne znaczenie w przypadkach realizacji systemów SSL przeznaczonych do pracy w czasie rzeczywistym.

2. LOKALIZACJA ŹRÓDŁA DŹWIĘKU W PRZESTRZENI

W celu uproszczenia algorytmu lokalizacji 3 – D źródła dźwięku proponuje się wykorzystanie typowych tzw. mikrofonów „soundfield”. Mikrofony takie produkowane są przez kilka firm, a eksperymenty wykonane przez autora zostały wykonane z wykorzystaniem „TetraMic Single Point Surround Sound Microphone” [5]. Standardowe zastosowanie mikrofonów „soundfield”, to rejestracja sygnału audio z przeznaczeniem do odtwarzania w wielokanałowych systemach dźwięku przestrzennego (systemy surround). Mikrofon „soundfield” może być wykorzystywany dwojako: mikrofon monofoniczny o dowolnie kształtowanej charakterystyce kierunkowej, matryca czterech mikrofonów umożliwiająca rejestrację sygnału w tzw. A-formacie. Możliwość wykorzystania drugiej opcji pokazano w dalszej części artykułu.

2.1 Formaty A i B

Mikrofon „soundfield” jest zbudowany z czterech kapsuł o charakterystyce kardoidalnej. Przykład jednego z nich pokazano na rysunku 3a. Kapsułki mikrofonów są zamontowane w konfiguracji takiej, jak ściany czworościanu foremnego, o orientacji pokazanej na rysunku 3b.



Rys. 3. Mikrofon „soudfield”: a) zdjęcie przykładowego mikrofonu „TetraMic Single Point Surround Sound Microphone”, b) orientacja w przestrzeni kapsułek mikrofonów.

Tak zbudowany mikrofon pozwala na zapisanie wszystkich potrzebnych informacji o polu dźwiękowym w danym punkcie przestrzeni. Na jego wyjściu otrzymuje się cztery niezależne sygnały, które mogą być dalej dowolnie przetwarzane. A-format jest sygnałem wprost z wyjść mikrofonów. B-format natomiast jest właściwym formatem studyjnym i jednocześnie bazowym dla systemu tzw. ambisonii wykorzystywanym w inżynierii dźwięku. Ten właśnie sygnał zostanie zastosowany w systemie lokalizacji kierunku źródła dźwięku.

Sygnał A-formatu zawiera cztery sygnały monofoniczne: LFU - lewy przedni górny, RBU - prawy tylni dolny, LBD - lewy tylni dolny i RFD - prawy przedni dolny, które pochodzą bezpośrednio z kapsułek mikrofonowych. Dla otrzymania sygnałów B-formatu sygnały A-formatu są poddawane przekształceniom arytmetycznym, dając sygnały X, Y, Z oraz W [6]:

$$W = v_{LFU} + v_{RBU} + v_{LBD} + v_{RFD} \quad (3)$$

$$X = v_{LFU} - v_{RBU} - v_{LBD} + v_{RFD} \quad (4)$$

$$Y = v_{LFU} - v_{RBU} + v_{LBD} - v_{RFD} \quad (5)$$

$$Z = v_{LFU} + v_{RBU} - v_{LBD} - v_{RFD} \quad (6)$$

gdzie v_{LFU} jest sygnałem wyjściowym kapsuły LFU, itd.

Wyliczone sygnały WXYZ mają interpretację następującą: W jest proporcjonalny do ciśnienia akustycznego w centralnym punkcie systemu mikrofonów z rysunku, natomiast sygnały X, Y, i Z są proporcjonalne do ciśnienia akustycznego na kierunkach odpowiednich osi 3-D układu współrzędnych: X – przód - tył, Y – lewo – prawo, Z – góra – dół.

Konwersja „A - B format”, to zamiana układu współrzędnych z układu czworościanu foremnego na XYZ. Dodatkowo uzyskuje się jeszcze informację o ciśnieniu akustycznym w środku układu współrzędnych XYZ.

Wzory od (1) do (4) definiują konwersję A na B format. Mając na uwadze geometrię czworościanu foremnego i miejsca ulokowania kapsułek, można zdefiniować zestaw wektorów określających ich lokalizację w przestrzeni:

$$\begin{aligned} \mathbf{u}_{LFU} &= \frac{1}{\sqrt{3}} \cdot [1 \ 1 \ 1]^T, \quad \mathbf{u}_{RBU} = \frac{1}{\sqrt{3}} \cdot [-1 \ -1 \ 1]^T \\ \mathbf{u}_{LBD} &= \frac{1}{\sqrt{3}} \cdot [-1 \ 1 \ -1]^T, \quad \mathbf{u}_{RFD} = \frac{1}{\sqrt{3}} \cdot [1 \ -1 \ -1]^T \end{aligned} \quad (7)$$

Przyjmując odległość każdej z kapsuł od środka układu współrzędnych XYZ równą r , miejsca geometryczne poszczególnych kapsuł są następujące:

$$\begin{aligned} \mathbf{x}_{LFU} &= r \cdot \mathbf{u}_{LFU}, & \mathbf{x}_{RBU} &= r \cdot \mathbf{u}_{RBU} \\ \mathbf{x}_{LBD} &= r \cdot \mathbf{u}_{LBD}, & \mathbf{x}_{RFD} &= r \cdot \mathbf{u}_{RFD} \end{aligned} \quad (8)$$

Uwzględniając geometrię układu mikrofonów, w punkcie \mathbf{x}_m ($m = LFU, RBU$ itd.) przestrzeni 3-D wystąpi ciśnienie akustyczne fali o pulsacji ω i amplitudzie A następującej wielkości [9]:

$$p_m = A \cdot e^{j(\omega t + \mathbf{k} \cdot \mathbf{x}_m)} \quad (9)$$

gdzie:

$$\mathbf{k} = k \cdot \mathbf{d} = k \cdot \begin{bmatrix} \cos(\theta) \cdot \cos(\varphi) \\ \sin(\theta) \cdot \cos(\varphi) \\ \sin(\varphi) \end{bmatrix} \quad (10)$$

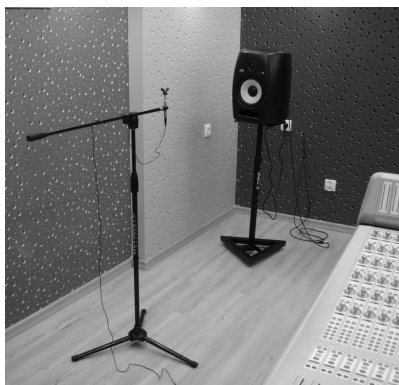
jest wektorem falowym, \mathbf{d} wektorem jednostkowym, wskazującym kierunek fali padającej, k - liczbą falową ($k=2\pi/\lambda$), θ - kątem elewacji, φ - kątem azymutu. Zakładając dodatkowo, że poszczególne kapsuły w mikrofonie „soundfield” mają kardoidalne charakterystyki kierunkowe i są typu ciśnieniowo – gradientowego [7], finalne otrzymuje się [9]:

$$X = \frac{W}{\sqrt{3}} \cdot \cos(\theta) \cdot \cos(\varphi), \quad Y = \frac{W}{\sqrt{3}} \cdot \sin(\theta) \cdot \cos(\varphi), \quad Z = \frac{W}{\sqrt{3}} \cdot \sin(\theta) \quad (11)$$

Z ostatnich wzorów wynika, że znając sygnał W oraz jedną parę z zestawu X, Y, Z z łatwością można wyznaczyć kierunek źródła dźwięku (azymut i elewację).

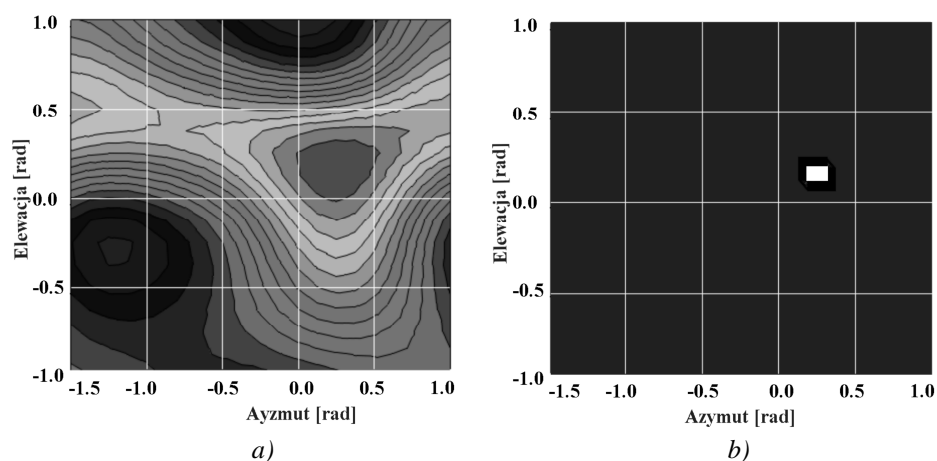
3. BADANIA EKSPERYMENTALNE

Testowanie proponowanego rozwiązania przestrzennego systemu SSL przeprowadzone w pomieszczeniu o rozmiarach 5x6x3 m i czasie pogłosu $T60$ [8] wynoszącym 0.8 s. Mikrofon „soundfield” umieszczony był ok. 1.5 m nad podłogą. Jako testowe źródło dźwięku wykorzystano głośnik odtwarzający sygnał z generatora szumu różowego. Odległość głośnika od mikrofonu ok. 1.5 m. Sygnały w formacie A nagrano cyfrowym rejestratorem Zaxcom DEVA 5.8.



Rys. 4. Pomieszczenie, w którym wykonano pomiary: Laboratorium Inżynierii Dźwięku i Ambiofonii, Zachodniopomorski Uniwersytet Technologiczny w Szczecinie

Dla różnych położeń źródła dźwięku nagrywano próbki o czasie trwania 10 s z częstotliwością próbkowania 96 kHz i z rozdzielczością 24 bity. W pierwszym etapie badań źródło dźwięku w czasie nagrania było nieruchome. Po dokonaniu konwersji z A na B format sygnały W, X, Y, Z podzielono na ramki o czasie trwania 250ms. Dla każdego zestawu ramek liczono kąt elewacji i azymutu, a następnie wynik uśredniano dla całkowitego czasu nagrania (10s). Wynik jednego z pomiarów pokazano na rysunku 5.



Rys. 5. Rezultat lokalizacji pojedynczego źródła dźwięku w 3-D układzie współrzędnych (widok z góry)

Na rysunku 5a pokazano bezpośrednio wyniki pomiarów. Rysunek 5b zawiera natomiast obraz z uwzględnieniem progowania [10]. Tę jedną z podstawowych technik przetwarzania

obrazu zastosowano w celu zwiększenia dokładności wyznaczenia kątów azymutu i elewacji.

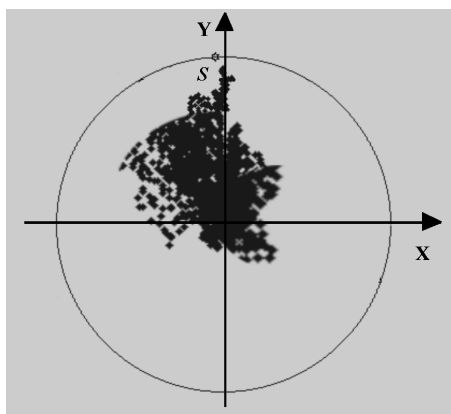
Wyniki eksperymentów dla kilku położeń źródła dźwięku względem mikrofonu przedstawiono w tabeli 1.

Tab. 1. Średnie błędy określania kierunku źródła dźwięku (A - azymut, E - elewacja)

Kierunek źródła dźwięku	$A=0^0$, $E=90^0$	$A=90^0$, $E=45^0$	$A=45^0$, $E=-45^0$	$A=-45^0$, $E=90^0$
Średni błąd kierunku	$\epsilon_A=4^0$, $\epsilon_E=5^0$	$\epsilon_A=6^0$, $\epsilon_E=5^0$	$\epsilon_A=3^0$, $\epsilon_E=4^0$	$\epsilon_A=4^0$, $\epsilon_E=6^0$

Testy, których wyniki pokazano w tabeli 1, przeprowadzono w ten sposób, że źródło dźwięku było nieruchome, natomiast zmieniano położenie mikrofonu. Dzięki temu możliwe było stosunkowo dokładne określenie rzeczywistego położenia źródła dźwięku względem mikrofonu. Trzeba wyraźnie także podkreślić, że błąd azymutu i elewacji uzyskano na podstawie obliczeń numerycznych, a nie analizy wykresów (rys. 5).

Wykorzystując mikrofon „soundfield” można także z powodzeniem śledzić w czasie rzeczywistym ruch źródła dźwięku. Na rysunku 6 pokazano wykres zmian położenia źródła w czasie ok. 300 s.



Rys. 6. Śledzenie położenia źródła dźwięku w płaszczyźnie poziomej

W tym przypadku (rys. 6), odległość poszczególnych punktów od środka układu współrzędnych jest proporcjonalna do ciśnienia akustycznego. Wykres wykonano dla przypadku przemieszczania się źródła w płaszczyźnie planarnej. Mikrofon „soundfield” umieszczony był w środku układu współrzędnych. Znacznik „S” oznacza bieżący kierunek źródła dźwięku. Zmiany odległości punktów od środka układu XY , świadczą o zmianie poziomu dźwięku źródła lub o zmianach jego odległości od mikrofonu. W przeprowadzonym doświadczeniu poziom głośności źródła był stały, natomiast jego odległość od mikrofonu ulegała zmianie.

4. WNIOSKI

W pracy zaproponowano system wyznaczania kierunku źródła dźwięku z wykorzystaniem mikrofonu „soundfield”. Rozwiązanie to może być przydatne np. w konstrukcji robotów, jako jeden z elementów rozpoznawania środowiska. Uzyskane wyniki badań symulacyjnych są bardzo zachęcające. Pokazane w tabeli 1 błędy świadczą o tym, że system ten lokalizuje kierunek źródła dźwięku nie gorzej niż ludzki słuch. Człowiek bowiem jest w stanie określić azymut i elewację na podstawie wrażeń słuchowych z dokładnością $6^{\circ} - 10^{\circ}$. Niestety, pojedynczy mikrofon „soundfield” nie umożliwia określenia odległości źródła dźwięku. Powodem tego jest mała odległość między kapsułami i zastosowanie np. metody TDOA (Time Delay of Arrival) [2] jest nieskuteczne. Natomiast, uzupełniając system o jeszcze jeden mikrofon „soundfield” (lub nawet pojedynczą kapsułę o charakterystyce dookólnej), będzie można określić odległość źródła dźwięku w przestrzeni 3-D. Warunkiem jest jednak, aby odległość dodatkowego mikrofonu od mikrofonu „soundfield” wynosiła nie mniej niż 15 cm.

5. BIBLIOGRAFIA

- [1] Li H., Yoshiara T., Zhao Q., Watanabe T., Huang J.: *A spatial sound localization system for mobile robots*, IEEE Volume, Issue, Trans. Instrum. and Meas., pp. 1-6, (2007)
- [2] Valian J.M.: *Robust Sound Source Localization Using a Microphone Array on a Mobile Robot*, IEEE/RSJ International Conference on Intelligent Robotics and Systems, pp. 1228 – 1233, (2003)
- [3] Martinson E.: *Hiding the Acoustic Signature of a Mobile Robot*, IEEE Int. Conf. on Intelligent Robots and Systems (IROS). San Diego, (2007)
- [4] Martinson E., Arkin R.C.: *Noise Maps for Acoustically Sensitive Navigation*, Proceedings of SPIE, vol 5609, (2004)
- [5] Core Sound: <http://www.core-sound.com/TetraMic/1.php>
- [6] Farrar K.: *Soundfield Microphone*, Wireless World, vol. 85, No. 1526, pp. 48 –50, (1979)
- [7] Jamroz A.: *The Design and Use of a Double Cardioid Stereophonic Microphone*, Journal of the Audio Engineering Society, vol. 8 No. 2, pp. 100 – 104, (1960)
- [8] Saberi K., Perrott D.R.: *Lateralization threshold obtained under conditions in which the precedence effect is assumed to operate*, J. Acoust. Soc. Am., vol. 87, pp. 1732-1737, (1990)
- [9] Nguyen N. T., Zgrzywa A., Czyżewski A. (Eds.), Kornatowski E.: *Advances in Multimedia and Network Information System Technologies*, Springer – Verlag, Berlin, 2010
- [10] Watkins C., Sadun A., Marenka S.: *Modern image processing: warping, morphing and classical techniques*, Academic Press Inc., 1993

Praca naukowa finansowana przez MNiSW ze środków na naukę w latach 2009 – 2010, jako projekt badawczy własny nr N N505 364 336.