ZESZYTY
NAUKOWE
POLITECHNIKI
ŚLĄSKIEJ

P. 3341/06

Jerzy KASPRZYK

KOMPUTEROWE WSPOMAGANIE IDENTYFIKACJI PROCESÓW

WYDAWNICTWO POLITECHNIKI ŚLĄSKIEJ

Gliwice 2006

POLITECHNIKA ŚLĄSKA
ZESZYTY NAUKOWE
NR 1720

P. 3341/06

Jerzy KASPRZYK

KOMPUTEROWE WSPOMAGANIE IDENTYFIKACJI PROCESÓW

GLIWICE 2006

Opiniodawcy
Prof. dr hab. inż. Adam GRZECH
Prof. dr hab. inż. Zbigniew NAHORSKI

Kolegium redakcyjne
REDAKTOR NACZELNY – Prof. dr hab. inż. Andrzej BUCHACZ
REDAKTOR DZIAŁU – Dr inż. Krzysztof SIMEK
SEKRETARZ REDAKCJI – Mgr Elżbieta LEŚKO

Wydano za zgodą
Rektora Politechniki Śląskiej

PL ISSN 0434-0760

© Copyright by
Wydawnictwo Politechniki Śląskiej
Gliwice 2006

Spis treści

Przedmowa ... **7**

Rozdział 1 Wprowadzenie ... **9**

1.1. Stan zagadnienia ... 10

1.2. Wkład autora ... 11

1.3. Układ pracy ... 12

1.4. Spis ważniejszych oznaczeń ... 15

Rozdział 2 Istota identyfikacji ... **19**

2.1. Eksperymentalna metoda badawcza ... 20
2.1.1. Podstawowe wymagania wobec identyfikacji ... 21
2.1.2. Procedura identyfikacji off-line ... 23
2.1.3. Potrzeba komputerowego wspomagania identyfikacji ... 25

2.2. Identyfikacja w ujęciu probabilistycznym ... 26
2.2.1. Wzór Bayesa ... 26
2.2.2. Funkcja wiarygodności ... 28
2.2.3. Eksperyment identyfikacyjny ... 30
2.2.4. Obiekt dynamiczny ... 31

2.3. Przykład obiektu elektroakustycznego ... 34
2.3.1. Układ kompensacji ATH ... 34
2.3.2. Opis obiektu i cel identyfikacji ... 36

Rozdział 3 Planowanie eksperymentu i przygotowanie danych ... **38**

3.1. Podstawowe wymagania ... 39

3.2. Komputerowe wspomaganie wykonania eksperymentu ... 41
3.2.1. Sprzężenie procesu z komputerem ... 43
3.2.2. Sygnał pobudzający ... 45
3.2.3. Próbkowanie ... 48
3.2.4. Testowanie poprawności uzyskanych danych ... 50

3.3. Wspomaganie w zakresie przygotowania danych ... 53
3.3.1. Edycja danych ... 53
3.3.2. Wstępne przetwarzanie danych ... 54
3.3.3. Usuwanie składowych deterministycznych ... 55

3.4. Planowanie eksperymentu dla identyfikacji obiektu elektroakustycznego ... 56
3.4.1. Sygnał pobudzający ... 57
3.4.2. Wstępne przetwarzanie danych ... 59

Rozdział 4 Struktura modelu ... **60**

4.1. Taksonomia modeli parametrycznych ... 61
4.1.1. Modele stochastyczne ... 61
4.1.2. Modele wejściowo-wyjściowe w postaci czarnej skrzynki ... 63
4.1.3. Modele liniowe i nieliniowe ... 65

4.2. Liniowy model wejściowo-wyjściowy ... 67
4.2.1. Typ i struktura liniowego modelu MISO ... 70
4.2.2. Wybór typu modelu liniowego ... 72

4.3. Struktura modelu dla obiektu elektroakustycznego ... 74

Rozdział 5 Procedura identyfikacji ... **78**

5.1. Zadania aproksymacji, falsyfikacji i walidacji ... 78
5.1.1. Kryterium aproksymacji a cel identyfikacji ... 79
5.1.2. Falsyfikacja modelu ... 80
5.1.3. Walidacja modelu ... 82
5.1.4. Walidacja w identyfikacji obiektu elektroakustycznego ... 84

5.2. Iteracyjny charakter procedury ... 86
5.2.1. Postać sformalizowana procedury ... 89
5.2.2. Potencjalne źródła błędów ... 93
5.2.3. Inne źródła informacji ... 96
5.2.4. Testy wspomagające wybór struktury ... 98

5.3. Wspomaganie w identyfikacji struktury modeli liniowych ... 98
5.3.1. Kryterium wyboru struktury ... 99
5.3.2. Algorytm poszukiwania dla modeli MISO ... 103
5.3.3. Przykład symulacyjny ... 113

5.4. Wspomaganie w testowaniu i analizie modeli liniowych ... 116
5.4.1. Testy wizualizacyjne ... 117
5.4.2. Badanie uwarunkowania numerycznego ... 119

Rozdział 6 Implementacja systemu komputerowego wspomagania identyfikacji ... **121**

6.1. Oprogramowanie systemu ... 123
6.1.1. Interfejs użytkownika ... 124
6.1.2. Elastyczność systemu a prostota użytkowania ... 126
6.1.3. Dwa paradygmaty programowania ... 128

6.2. System MULTI-EDIP ... 129
6.2.1. Historia systemu ... 130
6.2.2. Założenia projektowe ... 131
6.2.3. Użytkowanie systemu MULTI-EDIP ... 133

Rozdział 7 Wyniki identyfikacji obiektu elektroakustycznego ... **136**

7.1. Identyfikacja struktury ... 137

7.2. Identyfikacja w obecności zakłócenia ... 141

7.3. Podsumowanie ... 143

Rozdział 8 Wnioski końcowe ... **145**

Literatura ... **148**

Dodatek A Estymacja parametrów modeli liniowych ... **159**

A.1. Implementacja metody najmniejszych kwadratów ... 161

A.2. Implementacja rekurencyjnej metody błędu predykcji ... 162

Dodatek B Testy uwarunkowania macierzy korelacji ... **165**

Dodatek C Praca z systemem MULTI-EDIP ... **167**

Summary ... **174**

Contents

Preface ... **7**

Chapter 1. Introduction ... **9**

1.1. State of the art ... 10

1.2. Authors contribution ... 11

1.3. Lay-out of the book ... 12

1.4. Glossary of notations ... 15

Chapter 2. The essence of identification ... **19**

2.1. Experimental scientific method ... 20
2.1.1. Basic requirements ... 21
2.1.2. Off-line identification procedure ... 23
2.1.3. Necessity for computer aided identification ... 25

2.2. Identification in probabilistic formulation ... 26
2.2.1. Bayes' formula ... 26
2.2.2. Likelihood function ... 28
2.2.3. Identification experiment ... 30
2.2.4. Dynamical object ... 31

2.3. An example of electro-acoustic plant ... 34
2.3.1. ANC compensation system ... 34
2.3.2. Description of the plant and the goal of identification ... 36

Chapter 3. Planning the experiment and data preparation ... **38**

3.1. Basic requirements ... 39

3.2. Computer aided in experimentation ... 41
3.2.1. Proces-computer interface ... 43
3.2.2. Stimulating signal ... 45
3.2.3. Sampling ... 48
3.2.4. Testing the correctness of data ... 50

3.3. Computer aided data preparation ... 53
3.3.1. Data editing ... 53
3.3.2. Preprocessing ... 54
3.3.3. Elimination of deterministic components ... 55

3.4. Planning the experiment for the electro-acoustic plant ... 56
3.4.1. Stimulation ... 57
3.4.2. Data preprocessing ... 59

Chapter 4. Model structure ... **60**

4.1. Classification of parametric models ... 61
4.1.1. Stochastic models ... 61
4.1.2. Input-output black-box models ... 63
4.1.3. Linear and non-linear models ... 65

4.2. Linear input-outpu model ... 67
4.2.1. Type and structure of the linear MISO models ... 70
4.2.2. Model type selection ... 72

4.3. Model structure of the electro-acoustic plant ... 74

Chapter 5. The identification procedure **78**

5.1. Approximation, falsification and validation 78
5.1.1. Approximation criterion and the goal of identification 79
5.1.2. Model falsification 80
5.1.3. Model validation 82
5.1.4. Validation in electro-acoustic plant identification 84

5.2. Iterative procedure 86
5.2.1. Formal form of the procedure 89
5.2.2. Potential sources of faults 93
5.2.3. The other sources of information 96
5.2.4. Tests for model structure selection 98

5.3. Assistance in structure identification of linear models 98
5.3.1. Criterion for structure selection 99
5.3.2. Algorithm for MISO models 103
5.3.3. Simulation example 113

5.4. Assistance in testing and analysis of linear models 116
5.4.1. Visual tests 117
5.4.2. Testing numerical condition 119

Chapter 6. Implementation of computer aided identification **121**

6.1. Software 123
6.1.1. User interface 124
6.1.2. Flexibility versus simplicity 126
6.1.3. Two paradigmes of programming 128

6.2. System MULTI-EDIP 129
6.2.1. The history 130
6.2.2. Brief foredesign 131
6.2.3. Use of MULTI-EDIP 133

Chapter 7. The results of electro-acoustic plant identification **136**

7.1. Structure identification 137

7.2. Identification in the presence of disturbance 141

7.3. Summary 143

Chapter 8. Final conclusions **145**

References **148**

Appendix A. Parameter estimation for linear models **159**

A.1. Implementation of the least squares method 161

A.2. Implementation of the recursive prediction error method 162

Appendix B. Conditional tests for the correlation matrix **165**

Appendix C. Working with MULTI-EDIP **167**

Summary **174**

Przedmowa

Identyfikacja procesów, polegająca na wyznaczaniu modeli matematycznych opisujących ich właściwości na podstawie posiadanej *a priori* wiedzy oraz pomiarów zebranych w trakcie doświadczeń identyfikacyjnych, stała się przedmiotem interdyscyplinarnym, znajdującym zastosowanie zarówno w technice, jak i w wielu dziedzinach nietechnicznych, takich jak np. biologia, medycyna czy nauki społeczne. Stąd tak istotne znaczenie w pomyślnym jej przeprowadzeniu ma możliwość dostępu do odpowiedniego, i w jakimś stopniu inteligentnego, systemu komputerowego wspomagania, umożliwiającego korzystanie z osiągnięć teorii i praktyki identyfikacji osobom nie będącym ekspertami w tej dziedzinie. Celem tej pracy jest podsumowanie wieloletnich badań nad stworzeniem takiego systemu.

System do komputerowego wspomagania identyfikacji procesów rozwijany był w Instytucie Automatyki Politechniki Śląskiej od roku 1985, początkowo pod nazwą *EDIP*, później *MULTI-EDIP*. Doświadczenie zgromadzone w trakcie jego rozwijania i wdrażania u licznych użytkowników umożliwiło określenie założeń konstrukcyjnych oraz zaproponowanie mechanizmów wspomagania w prowadzeniu identyfikacji, dzięki którym uzyskano system będący zintegrowanym środowiskiem programowo-sprzętowym. Bardzo ważną własnością tego systemu jest to, że użytkownik nie musi posiadać żadnej wiedzy i żadnych umiejętności z dziedziny programowania komputerów, gdyż użytkuje go tak jak przyrząd mikrokomputerowy.

Zebrane doświadczenia pozwoliły na sformułowanie pewnych bardziej ogólnych wniosków dotyczących możliwości i zakresu wspomagania użytkownika w procedurze identyfikacji oraz zasad poprawnego jej wykonywania. Stąd w pracy przedstawiona została analiza zadań dla komputerowego systemu wspomagania identyfikacji, a w szczególności dyskusja praktycznych rozwiązań w implementacji takiego systemu oraz ograniczeń, na jakie można napotkać podczas jego stosowania.

Obecny kształt praca uzyskała także dzięki Recenzentom: Panu Profesorowi Zbigniewowi Nahorskiemu z Instytutu Badań Systemowych PAN oraz Panu Profesorowi Adamowi Grzechowi z Politechniki Wrocławskiej, którym chciałbym tą drogą złożyć serdeczne podziękowania za wnikliwość i wiele cennych uwag.

Szczególnie gorąco pragnę podziękować również Panu Profesorowi Antoniemu Niederlińskiemu za inspirację do napisania tej pracy oraz stałe wsparcie w trakcie jej pisania, a także Dr. hab. Jarosławowi Figwerowi za wiele lat wspólnej pracy nad systemem *MULTI-*

EDIP oraz istotne komentarze podczas pisania. Podobnie dziękuję pozostałym Kolegom z Zakładu Komputerowych Systemów Sterowania w Instytucie Automatyki.

Nie bez znaczenia było dla mnie również wsparcie i wyrozumiałość członków mojej Rodziny, którym tą drogą chciałbym złożyć moje podziękowanie.

Praca ta powstała dzięki częściowemu finansowaniu z grantu Ministerstwa Edukacji i Nauki o numerze 3 T11A 012 29.

Jerzy Kasprzyk

Gliwice, maj 2006

Rozdział 1

Wprowadzenie

Jedną z podstawowych czynności w nauce i technice jest modelowanie, a w szczególności *modelowanie matematyczne*. Współczesna nauka dysponuje bogatym wachlarzem metod umożliwiających wyznaczanie modeli matematycznych różnych procesów. Wyróżnić tu można dwa podstawowe podejścia do problemu:

- fenomenologiczne, oparte na analizie zjawisk fizycznych, chemicznych lub innych, na podstawie których tworzy się model matematyczny opisujący własności modelowanego procesu,
- eksperymentalne, które obejmuje się nazwą *identyfikacji procesów*.

Identyfikacja procesów polega na tworzeniu *modeli matematycznych* (z wybranej klasy modeli) opisujących właściwości procesów na podstawie posiadanej wiedzy *a priori* oraz *pomiarów* sygnałów zebranych w trakcie – jeżeli to możliwe – specjalnie zaplanowanych *doświadczeń identyfikacyjnych* [107].

Identyfikację procesów można więc traktować jako *uogólnioną technikę pomiarową*, umożliwiającą „pomiar dynamiki w sposób pośredni" [101]. W tym celu potrzebny jest *sprzęt komputerowy* do zbierania danych i ich przetwarzania oraz wysokiej jakości *oprogramowanie*, oparte na złożonych i skomplikowanych *podstawach matematycznych*, umożliwiające przetworzenie zbioru danych pomiarowych w model matematyczny, stanowiący zwięzłą formę opisu najważniejszych właściwości procesu. *Identyfikacja* (wyznaczanie modelu na podstawie przebiegów sygnałów wejściowych i wyjściowych) może być uważana za problem odwrotny do *analizy* (wyznaczanie wartości wyjścia na podstawie znajomości wejścia i modelu), stąd jest ona zagadnieniem trudniejszym matematycznie.

Tematem pracy jest analiza możliwości komputerowego wspomagania w zakresie identyfikacji procesów, w szczególności dyskusja zadań i praktycznych możliwości implementacji takiego systemu do identyfikacji dynamicznych modeli liniowych, w tym także ograniczeń, na jakie można napotkać w trakcie stosowania systemu.

1.1. Stan zagadnienia

Identyfikacja procesów ma już długą historię i dostępna jest bogata literatura w tym zakresie, np. znane podręczniki [27, 46, 62, 79, 86, 107, 115, 132]. Podręczniki te zawierają głównie opisy różnych metod identyfikacji oraz dyskusję ich teoretycznych własności. Aby móc korzystać z proponowanych metod, potrzebny jest jeszcze, oprócz dużej wiedzy w tym zakresie, dostęp do odpowiedniego *oprogramowania.*

Jednak praktyczne stosowanie identyfikacji wymaga także pewnych umiejętności. Wyniki identyfikacji zależą bowiem nie tylko od zastosowanych procedur numerycznych, które powinny być zarówno *dokładne* (tj. umożliwiać wyznaczenie modeli z odpowiednią precyzją), jak i *odporne* (tj. mało wrażliwe na losowe zmiany danych stosowanych do identyfikacji). W nie mniejszym stopniu wyniki te zależą także od *sposobu* prowadzenia identyfikacji, a więc od doświadczenia i umiejętności osoby, która tę identyfikację wykonuje. Umiejętności te są potrzebne zarówno na etapie planowania eksperymentu identyfikacyjnego, modelowania polegającego na przyjęciu pewnych założeń dotyczących klasy czy struktury modelu, jak i przy wyborze oraz parametryzacji odpowiedniej procedury numerycznej, umożliwiającej estymację modelu, czy wreszcie na etapie analizowania i testowania wyników identyfikacji. Identyfikacja modeli jest więc swoistym połączeniem trudnych metod matematycznych, różnorodnych technik programowania oraz doświadczenia praktycznego.

Jednocześnie jest ona techniką o bardzo rozległych, interdyscyplinarnych zastosowaniach. Potrzeba wyznaczania modeli matematycznych na podstawie pomiarów występuje bowiem w tak różnorodnych dziedzinach, jak np.:

- technika, m.in. aeronautyka i astronautyka, akustyka i wibroakustyka, automatyzacja, bioinżynieria, inżynieria chemiczna, technika regulacji, elektronika, mechanika, diagnostyka techniczna i telekomunikacja;
- nauki podstawowe, m.in. fizyka i chemia;
- nauki przyrodnicze i geograficzne, m.in. biologia, medycyna, sejsmografia i hydrologia;
- nauki społeczne, m.in. ekonomia i socjologia.

Tak więc użytkownik oprogramowania do identyfikacji powinien być *ekspertem* zarówno w zakresie dotyczącej go wiedzy dziedzinowej, jak i w zakresie techniki identyfikacji. Stąd tak duże znaczenie ma możliwość dostępu do odpowiedniego, i w jakimś stopniu *inteligentnego*, oprogramowania wspomagającego, które umożliwiłoby korzystanie z osiągnięć teorii identyfikacji osobom nie będącym ekspertami w tej dziedzinie. Potrzebę komputerowego wspomagania identyfikacji omówiono bliżej w [99], natomiast w [102] przeprowadzono

dyskusję, co można uważać za oprogramowanie inteligentne. Domyślnie bowiem, na podstawie analizy tego, co się na ogół pisze na ten temat, przyjmuje się, że jakiś algorytm czy program do sterowania lub identyfikacji jest uznawany za *inteligentny*, jeśli zawiera w sobie paradygmat (lub jest oparty na paradygmacie), który przypomina sposób wnioskowania człowieka, por. [112].

Jednak problem, w jakim stopniu identyfikację można przeprowadzić w sposób automatyczny, właściwie do tej pory nie znalazł precyzyjnego rozwiązania, aczkolwiek w literaturze prezentowanych jest wiele interaktywnych pakietów do identyfikacji, jak np. *ESPION* [45], *SEXI* [37], *EFPI* [97], *MULTI-EDIP* [64], *ISID* [141], *ADAPTX* [74]. Z kolei w [17] przedstawiono dyskusję perspektyw, jakie stwarza identyfikacja interaktywna, oraz możliwości powstania pułapek. W [80] natomiast omówiono wymagania dydaktyczne, jakie powinien spełniać interfejs oprogramowania dla celów komputerowego wspomagania identyfikacji. W przypadku najbardziej popularnego w środowisku naukowym systemu *MATLAB* wraz z przybornikiem *Identification Toolbox* [81] zakres wspomagania użytkownika sprowadza się do dostarczenia zestawu funkcji, które mogą być wykorzystane w trakcie identyfikacji. Nie umożliwia on jednak prowadzenia użytkownika przez kolejne etapy identyfikacji i wymaga od niego sporej wiedzy w tym zakresie.

1.2. Wkład autora

W pracy przedstawiono koncepcję systemu komputerowego wspomagania identyfikacji, stanowiącą pewne podsumowanie dotychczasowych wyników autora w tym zakresie. Na tle ogólnych rozważań na temat celów i metod identyfikacji przeprowadzono dyskusję, w jakim stopniu procedura identyfikacji może być wykonana automatycznie, bez udziału użytkownika systemu, a w jakim zakresie musi on przejąć odpowiedzialność za jej prowadzenie. W efekcie przedstawiono całą procedurę identyfikacji w postaci sformalizowanej, która może stanowić podstawę do projektowania komputerowego systemu wspomagania. Omówiono, w jaki sposób użytkownik takiego systemu może być wspomagany w trakcie przechodzenia przez kolejne etapy tej procedury, poczynając od planowania eksperymentu po testowanie otrzymanego modelu.

Ponieważ w zastosowaniach zdecydowanie największą popularnością cieszą się liniowe modele stacjonarne (ang. *Linear Time Invariant, LTI*), stąd główną uwagę skupiono na tej klasie modeli. Szczególne znaczenie ma tu zaproponowanie algorytmu identyfikacji struktury dla parametrycznych modeli *LTI* obiektów wielowejściowych *MISO* (*Multi-Input-Single-Output*),

jako że jednym z często spotykanych problemów w praktyce jest brak możliwości określenia *a priori* struktury takiego modelu.

Przedstawione w pracy propozycje komputerowego wspomagania identyfikacji zostały zaimplementowane w systemie *MULTI-EDIP*, nad którym prowadzono prace w Instytucie Automatyki przez ostatnich kilkanaście lat, por. 6.2. Udział autora w pracach nad tym systemem obejmował, oprócz tworzenia koncepcji całego systemu, przede wszystkim opracowanie tej części, która związana jest z identyfikacją modeli parametrycznych, a także z wykonywaniem eksperymentu identyfikacyjnego na rzeczywistym obiekcie i przygotowaniem danych. Implementacja wielu znanych z literatury algorytmów wymagała dodatkowo opracowania i testowania odpowiednich procedur odpornych numerycznie, których szczegółowe omówienie tu pominięto. Przy okazji wprowadzono pewne modyfikacje algorytmów, które w istotny sposób przyspieszały działanie systemu, jak np. w przypadku metody najmniejszych kwadratów, lub poprawiały dokładność wyników, jak np. odpowiednia implementacja metody *RPE*, por. dodatek A. Problemem, który na ogół pomijany jest w literaturze, jest zagadnienie badania uwarunkowania równań estymatora. Stąd zaproponowano testy pomocnicze umożliwiające określenie przyczyn słabego uwarunkowania, które przedstawiono w dodatku B. Zaproponowanych zostało także kilka prostych testów wspomagających użytkownika na etapie przygotowania danych, por. 3.2.4, oraz wstępnej oceny stacjonarności procesu, por. 5.2.4.

Przy wyborze metod do implementacji w systemie *MULTI-EDIP* autor kierował się zasadą, że metody te powinny być nie tylko dobrze uzasadnione teoretycznie, ale także zrozumiałe intuicyjnie dla użytkownika nie będącego specjalistą w dziedzinie identyfikacji czy w zakresie statystyki. Za pewien wkład autora można tu także uznać zaproponowanie takiej realizacji interfejsu użytkownika, która stanowiłaby pewien kompromis między elastycznością a prostotą użytkowania systemu, por. 6.1.2.

1.3. Układ pracy

Wychodząc od przedstawionej przez Poppera w [116] istoty eksperymentalnej metody badawczej oraz od probabilistycznego spojrzenia na problem wyznaczania modeli na drodze identyfikacji, w rozdziale drugim omówiono podstawowe wymagania wobec identyfikacji prowadzonej w trybie *off-line*. Przedstawiono ogólną procedurę identyfikacji, zwracając uwagę na konieczność komputerowego wspomagania w trakcie jej realizacji. W rozdziale tym omówiono także przykład rzeczywistego obiektu elektroakustycznego, którego identyfikacja jest konieczna w celu prawidłowej parametryzacji układu aktywnego tłumienia hałasu. Obiekty tego typu są szczególnie trudne do identyfikacji, gdyż cechują się bardzo złożoną dynamiką,

której nie można modelować za pomocą filtrów dolnoprzepustowych, jak w przypadku typowych procesów przemysłowych. Przykład ten będzie się przewijał przez kolejne rozdziały w celu praktycznej ilustracji rozważań teoretycznych.

Kluczowe znaczenie dla uzyskania sukcesu ma prawidłowe przeprowadzenie doświadczenia identyfikacyjnego oraz odpowiednie przygotowanie danych, stąd rozdział trzeci zawiera omówienie podstawowych wymagań wobec eksperymentu, warunkujących poprawne wnioskowanie na podstawie uzyskanych danych. Przedstawiono aspekty techniczne sprzężenia komputera z procesem oraz możliwości wspomagania użytkownika w tym zakresie. Poruszono również problem wspomagania w przygotowaniu danych i zaproponowano kilka prostych testów badania uwarunkowania danych, których zastosowanie może już na wstępnym etapie identyfikacji pozwolić na uniknięcie niepotrzebnych błędów.

Istotne jest także przyjęcie odpowiedniej struktury modelu, która powinna być odpowiednia dla opisu własności identyfikowanego procesu, ale także uwzględniać cel identyfikacji. W rozdziale czwartym przedstawiono ogólną taksonomię różnych struktur stochastycznych modeli parametrycznych, a w szczególności skupiono się na problemie doboru struktury dla liniowych modeli wejściowo-wyjściowych.

W rozdziale piątym omówiono podstawowe zadania cząstkowe wykonywane w trakcie procedury identyfikacji, takie jak określenie kryterium aproksymacji oraz zadania falsyfikacji i walidacji modelu. Na tej podstawie przedstawiona została procedura identyfikacji w postaci sformalizowanej z uwzględnieniem tych elementów, które mogą być automatycznie wykonywane przez komputer. Omówiono także możliwe źródła błędów oraz zaproponowano kilka prostych testów wspomagających określenie struktury modelu. Następnie, zawężając obszar zainteresowań do klasy modeli najczęściej stosowanych, skupiono się na problemie identyfikacji modeli *LTI* typu czarna skrzynka. Przedyskutowano możliwości komputerowego wspomagania w zakresie określenia struktury takich modeli, w tym problem kryterium wyboru struktury. Zaproponowano procedurę automatycznego wyznaczania struktury modelu *MISO* minimalizującej wybrane kryterium w połączeniu z zadaniem estymacji jego parametrów. Zagadnienie to jest szczególnie ważne w sytuacji komputerowego wspomagania identyfikacji procesu, którego struktura jest użytkownikowi nieznana albo trudna do określenia *a priori*. Przedstawiono także propozycje testów wspomagających walidację i falsyfikację modelu, w szczególności analizę numeryczną uzyskanych wyników.

Rozdziały 3 do 5 zostały napisane w ten sposób, że najpierw wychodzi się od ogólnych rozważań wynikających z zasad wnioskowania statystycznego, po czym bardziej szczegółowo rozpatruje się klasę modeli liniowych wejściowo-wyjściowych, a całość ilustrowana jest przykładem zadania identyfikacji wspomnianego obiektu elektroakustycznego.

W rozdziale szóstym omówiono wymagania, jakim powinno sprostać oprogramowanie systemu wspomagającego użytkownika w prowadzeniu identyfikacji, w szczególności w zakresie interfejsu użytkownika. Przeprowadzono dyskusję możliwych rozwiązań i podkreślono konieczność pewnego kompromisu między elastycznością systemu a prostotą użytkowania. Uwieńczeniem wieloletnich prac nad stworzeniem systemu wspomagającego użytkownika w identyfikacji procesów jest pakiet programowy o nazwie *MULTI-EDIP*, który został zaprojektowany tak, aby jak najlepiej spełniał omawiane wymagania.

W rozdziale siódmym przedstawiono wyniki wykorzystania systemu *MULTI-EDIP* do identyfikacji modelu przykładowego obiektu elektroakustycznego.

Rozdział ósmy zawiera podsumowanie rozważań i wnioski, w tym omówienie perspektyw i możliwych zagrożeń w rozwijaniu i stosowaniu systemów komputerowego wspomagania identyfikacji.

W dodatku A przedstawiono krótki opis implementacji klasycznych metod estymacji parametrów modeli z zaproponowanymi modyfikacjami, natomiast dodatek B zawiera szczegóły dotyczące testów uwarunkowania danych. W dodatku C z kolei pokazano, w jaki sposób użytkownik współpracuje z systemem *MULTI-EDIP* wykorzystując jego interfejs graficzny.

Praca jest ukierunkowana przede wszystkim na praktyczne aspekty identyfikacji procesów, w szczególności związane z dyskusją możliwości i zakresu komputerowego wspomagania identyfikacji oraz implementacją odpowiednich metod i algorytmów. Stąd prowadzone rozważania w dużej mierze oparte są na heurystyce wynikającej z doświadczenia autora. Wynika to ze specyfiki identyfikacji, która jest nie tylko nauką, ale także swego rodzaju sztuką [80]. W przypadku projektowania systemu komputerowego wspomagania identyfikacji problem polega na tym, jak to doświadczenie i wiedzę zapisać w systemie, aby umożliwić użytkownikowi, nawet nie będącemu specjalistą w tym zakresie, pomyślne przeprowadzenie identyfikacji. W pracy pominięto analizę zagadnień teoretycznych, takich jak np. własności asymptotyczne typu zbieżność, nieobciążoność czy zgodność, które można znaleźć w wielu publikacjach dotyczących identyfikacji.

1.4. Spis ważniejszych oznaczeń

Oznaczenia ogólne

i – czas dyskretny, numer próbki

$x(i)$ – wartość x w chwili i

x_k – k-ty element x, jeśli x jest wektorem

$\mathbf{x}$ – zaobserwowana wartość zmiennej x (realizacja zmiennej losowej $\mathcal{X}$)

$\hat{x}$ – obliczona ocena zmiennej x

$p(x)$ – gęstość prawdopodobieństwa zmiennej losowej $\mathcal{X}$ w punkcie x

$p(x \mid y)$ – warunkowa gęstość prawdopodobieństwa

z^{-1} – operator opóźnienia

$A(z^{-1})$ – wielomian operatora z^{-1}

ln – logarytm naturalny

ϕ – zbiór pusty

$\leftarrow$ oznacza wyznaczenie wartości lub instrukcję przypisania

$\rightarrow$ oznacza, że to, co występuje po lewej stronie znaku, determinuje to, co jest po prawej

Literą pogrubioną oznaczana jest wartość rzeczywista zmiennej (realizacja zmiennej losowej). Daszkiem oznaczana jest obliczona ocena zmiennej.

Zmienne i zbiory

$\mathcal{A}$ – kryterium aproksymacji

d – czas opóźnienia (dyskretny)

$d_{\min}$, $d_{\max}$ – zakres poszukiwanych czasów opóźnienia

dA, dB, dC, dD, dF – stopnie wielomianów $A(z^{-1})$, $B(z^{-1})$ itd.

e – biały szum

$\mathcal{F}$ – postać modelu charakteryzująca strukturę $\mathcal{M}$ (np. równanie różnicowe wejściowo–wyjściowe), dla modelu liniowego $\mathcal{F}$ oznacza typ modelu

f – częstotliwość [Hz]

f_p – częstotliwość próbkowania, $f_p = 1/\tau_p$

$G(z^{-1})$ – transmitancja toru sterowania

$H(z^{-1})$ – transmitancja toru zakłócenia

$L(x)$ – funkcja wiarygodności zmiennej x

M – model obiektu $\mathbf{S}$, $M = (\mathcal{F}, \mu, \theta)$

$\mathbf{M}$ – model równoważny obiektowi $\mathbf{S}$ ze względu na cel Q

$\hat{M}$ - model optymalny (wybrany odpowiednio do kryterium aproksymacji $\mathcal{A}$)

$\mathcal{M}$ – struktura modelu (zbiór wszystkich modeli próbnych), $\mathcal{M} = (\mathcal{F}, \mu)$

$\mathcal{M}_{max}$ – najogólniejsza struktura modelu

m – liczba parametrów, wymiar wektora θ

N – długość próbki danych

$\mathcal{N}$ – zbiór modeli niesfalsyfikowanych, $\mathcal{N}(\gamma \mid \mathbf{U}_N, \mathbf{Y}_N) = \{M \mid \Gamma(M \mid \mathbf{U}_N, \mathbf{Y}_N) \le \gamma\}$

n – wymiar μ lub stopień wielomianów w procedurze identyfikacji

n_{min}, n_{max} - zakres poszukiwanych stopni wielomianów

O – otoczenie procesu P i eksperymentatora R

P – proces

$\mathbf{P}$ – proces rzeczywisty

Q – cel identyfikacji, zbiór modeli odpowiednich do celu, $Q(\delta \mid \mathbf{M}) = \{M \mid \Delta(M \mid \mathbf{M}) \le \delta\}$

Q – kryterium wyboru struktury modelu

q – liczba wejść w obiekcie (wymiar wektora u)

R – źródło pobudzenia, sterowanie procesem

$\mathbf{R}$ – rzeczywiste źródło pobudzenia w czasie $\mathbf{T}$ (eksperymentator)

$\mathbf{R}_{max}$ – najobszerniejszy eksperymentator

r – parametr charakteryzujący R

S – obiekt identyfikacji, $S = (O, P)$;

$\mathbf{S}$ – rzeczywisty obiekt identyfikacji

T – czas stosowania modelu

$\mathbf{T}$ – czas wykonywania eksperymentu identyfikacyjnego

t – czas (ciągły)

u – zmienna wejściowa

$\mathbf{u}$ – realizacja zmiennej wejściowej (pobudzenie w czasie eksperymentu)

U_N – ciąg wartości wejść $U_N \equiv \{u(i) \mid i = 1, 2, \ldots, N\}$

V – funkcja strat

x - wektor stanu

y – zmienna wyjściowa

$\mathbf{y}$ – realizacja zmiennej wyjściowej (odpowiedź w czasie eksperymentu)

Y_N – ciąg wartości wyjść, $Y_N \equiv \{y(i) \mid i = 1, 2, \ldots, N\}$

Γ – miara odległości między modelem M a danymi $(\mathbf{U}_N, \mathbf{Y}_N)$

γ – graniczna wartość Γ dla akceptacji modelu

Δ – miara odległości między modelem M a modelem $\mathbf{M}$ równoważnym obiektowi $\mathbf{S}$ ze względu na cel Q

δ – granica dokładności dla określenia zbioru modeli odpowiednich do celu Q

ε – błąd identyfikacji (zazwyczaj jest to błąd predykcji)

θ – zbiór parametrów modelu M

$\hat{\theta}$ – ocena θ uzyskana dla kryterium $\mathcal{A}$

μ – zbiór wskaźników struktury $\mathcal{M}$

$\rho_{yy}(l)$ – wartość funkcji autokorelacji zmiennej y dla przesunięcia l

$\rho_{xy}(l)$ – wartość funkcji korelacji wzajemnej zmiennych x i y dla przesunięcia l

τ_p – okres próbkowania

Ω – zbiór zdarzeń elementarnych

ϖ – zdarzenie elementarne

Ω – częstotliwość podstawowa (bin częstotliwości), $\Omega = 2\pi/N$

ω – częstotliwość kołowa [rad/s], $\omega = 2\pi f$

$(\mathbf{R}, \mathbf{S})$ – system eksperymentalny (źródło danych)

$(\mathbf{U}_N, \mathbf{Y}_N)$ – zbiór danych

Skróty

AIC (*Akaike Information Criterion*) – kryterium informacyjne Akaike

ARMA (*AutoRegressive Moving Average*) – model autoregresji średniej ruchomej

ARMAX (*AutoRegressive Moving Average eXternal*) - model autoregresji średniej ruchomej z zewnętrznym pobudzeniem

ARX (*AutoRegressive eXternal*) – model autoregresji z zewnętrznym pobudzeniem

BIC (*Bayes Information Criterion*) – kryterium informacyjne Bayesa

BJ (*Box-Jenkins*) – model Boxa-Jenkinsa

DAF (*Disturbance Adjusted Filtering*) – filtracja dostosowana do zakłócenia

ETFE (*Empirical Transfer Function Estimator*) – doświadczalny estymator funkcji przejścia

FIR (*Finite Impulse Response*) – model skończonej odpowiedzi impulsowej

FPE (Final Prediction Error) – kryterium błędu predykcji

FOE (*Final Output Error*) – kryterium błędu wyjścia

IV (*Instrumental Variable*) – metoda zmiennej pomocniczej (instrumentalnej)

LS (*Least Squares*) – metoda najmniejszych kwadratów

LTI (*Linear Time Invariant*) – model liniowy stacjonarny

MA (*Moving Average*) – model średniej ruchomej

MAP (*Maximum A Posteriori*) – metoda maksimum gęstości prawdopodobieństwa *a posteriori*

MIMO (*Multi Input Multi Output*) – obiekt o wielu wejściach i wielu wyjściach

MISO (*Multi Input Single Output*) – obiekt o wielu wejściach i jednym wyjściu

ML (*Maximum Likelihood*) – metoda największej wiarygodności

NSR (*Noise to Signal Ratio*) – stosunek szumu do sygnału użytecznego
OE (*Output Error*) – model błędu wyjścia
OS – odchylenie standardowe
PE (*Persistently Exciting*) – sygnał nieustannie pobudzający
PEM (*Prediction Error Methods*) – metody błędu predykcji
RPE (*Recursive Prediction Error*) – rekurencyjna metoda błędu predykcji
SISO (*Single Input Single Output*) – obiekt o jednym wejściu i jednym wyjściu

Rozdział 2
Istota identyfikacji

Rozdział ten jest poświęcony przedstawieniu ogólnych idei oraz zasad tworzenia modelu na podstawie eksperymentu identyfikacyjnego. Materiał prezentowany jest głównie na podstawie literatury przedmiotu, jednak w celu dyskusji możliwości komputerowego wspomagania identyfikacji oraz wymagań wobec projektowanego systemu konieczne jest uściślenie podstawowych pojęć i idei.

Stąd najpierw omówione zostaną podstawowe wymagania wobec metod badawczych opartych na eksperymencie, do których należy także identyfikacja, oraz potrzeba komputerowego wspomagania w tym zakresie. Następnie istota identyfikacji będzie przedstawiona w ujęciu *probabilistycznym*, opartym na wzorze Bayesa, w którym za zmienne losowe uważane są nie tylko zmienne w czasie sygnały (dane), ale także identyfikowany proces. Podejście takie nie jest zbyt popularne w literaturze dotyczącej identyfikacji procesów, aczkolwiek zalecane przez niektórych autorów, por. np. [16, 114, 17]. Ma ono przede wszystkim znaczenie *teoretyczne* i *filozoficzne*, pozwala bowiem na przedstawienie spójnej teorii identyfikacji procesów opartej na wnioskowaniu statystycznym, którego celem jest skorygowanie przyjętego *a priori subiektywnego* rozkładu prawdopodobieństwa poprzez wykorzystanie *obiektywnych* danych eksperymentalnych, por. [114]. Idea przetwarzania gęstości prawdopodobieństwa *a priori* w celu uzyskania prawdopodobieństwa *a posteriori* dobrze pasuje do estymacji rekurencyjnej. Tu podejście bayesowskie zostanie wykorzystane głównie do analizy warunków dla prawidłowego wykonania eksperymentu identyfikacyjnego oraz zasad wnioskowania na podstawie uzyskanych danych.

Jednak podejście bayesowskie, aczkolwiek płodne koncepcyjnie, jest trudne do praktycznego wykorzystania, gdy mamy do czynienia ze zmienną losową wielowymiarową, gdyż estymator ten jest zbyt wymagający obliczeniowo. Poza tym w praktycznych zastosowaniach pojawia się problem, w jaki sposób należy interpretować prawdopodobieństwo dla poszczególnych zmiennych (sygnałów) czy modelowanego procesu, a co za tym idzie, jak wyznaczyć odpowiednie funkcje gęstości prawdopodobieństwa. Zakładając ergodyczność (co niekoniecznie musi być prawdą) trzeba by wykonać bardzo długie doświadczenia, aby móc powiedzieć coś rozsądnie na temat wymaganych rozkładów prawdopodobieństwa. Zwykle nie jest to możliwe, a często ekonomicznie nieuzasadnione. Praktyczna trudność w wyznaczeniu

gęstości prawdopodobieństwa dla poszczególnych zmiennych występuje niezależnie od tego, jak definiowane jest samo prawdopodobieństwo - według klasycznej koncepcji Laplace'a, czy według częstościowej (statystycznej), zaproponowanej przez von Misesa, czy wreszcie zgodnie z aksjomatyką sformułowaną przez Kołmogorowa.

Ponieważ w pracy główny nacisk został położony na *praktyczne* aspekty identyfikacji, więc podstawowym sposobem opisu stosowanym w dalszej jej części będzie opis w postaci modelu *stochastycznego*, w którym niepewność jest przedstawiana za pomocą tzw. innowacji, będących elementarnymi zmiennymi losowymi o zadanym rozkładzie prawdopodobieństwa. Podejście takie jest szczególnie wygodne dla celów predykcji, sterowania czy symulacji i stąd znacznie popularniejsze w literaturze. W takim przypadku stosuje się estymatory, które można uważać za uproszczone wersje estymatora bayesowskiego, por. [107]. Szczególnie popularny jest estymator największej wiarygodności (*Maximum Likelihood*, *ML*) ze względu na dobre własności asymptotyczne i umiarkowane wymagania obliczeniowe. Podejście to będzie preferowane w dalszych rozważaniach także dlatego, że jest intuicyjnie atrakcyjne.

Istotę identyfikacji można również omawiać nie korzystając w ogóle z pojęć rachunku prawdopodobieństwa, por. np. [108], jednak wtedy trudno jest dokonać analizy ogólnych warunków poprawnej identyfikacji.

Na zakończenie rozdziału przedstawiono obiekt elektroakustyczny, który w dalszej dyskusji będzie stanowić przykład wiodący dla prezentacji praktycznego zastosowania omawianych zagadnień teoretycznych. Identyfikacja modeli obiektów elektroakustycznych jest jednym z typowych zadań przy projektowaniu układów aktywnego tłumienia hałasu (ATH). Tutaj omawiana będzie identyfikacja modeli transmitancji torów wtórnych i toru sprzężenia akustycznego potrzebnych do prawidłowej parametryzacji trójwymiarowego układu ATH dla tworzenia lokalnych stref ciszy w pomieszczeniu. Tego typu obiekty znacząco różnią się od typowych procesów przemysłowych. Przede wszystkim nie przenoszą one składowej stałej, a ich dynamika jest bardzo skomplikowana (szczególnie ze względu na własności akustyczne w pomieszczeniu z pogłosem). Obiekt taki nie może więc być modelowany za pomocą filtru dolnoprzepustowego niskiego rzędu, jak to zwykle bywa w przypadku obiektów ciągłych sterowanych cyfrowo, por. [57].

2.1. Eksperymentalna metoda badawcza

Na zadanie tworzenia modelu na drodze identyfikacji można spojrzeć z punktu widzenia bardziej ogólnego, jako na jedno z zagadnień stanowiących przedmiot zainteresowania nauk empirycznych, którym jest próba opisu jakiegoś fragmentu rzeczywistości za pomocą metod

matematycznych. Stąd definiując zasady prawidłowego modelowania i identyfikacji można korzystać z dorobku filozofii nauki. Zgodnie z *teorią metody naukowej* przedstawioną przez Poppera [116], warunkiem uznania prawomocności jakiegoś zdania naukowego w przypadku nauk empirycznych jest nie tyle *weryfikacja* jego prawdziwości, co *falsyfikacja*.

Logika eksperymentalnej metody badawczej opiera się bowiem na takich podstawowych pojęciach, jak: hipoteza, obserwacja i falsyfikacja. W metodzie tej można wyróżnić trzy kolejne etapy:

- postawienie *hipotezy*;
- wykonanie eksperymentu (*obserwacja*);
- testowanie, czy uzyskane dane pozwalają na odrzucenie możliwych do wywnioskowania skutków hipotezy (*falsyfikacja*).

Etapy te powtarza się wielokrotnie. Jeśli bowiem obserwacje pozwalają na odrzucenie hipotezy, to hipotezę należy poprawić, wykonać odpowiedni dla niej nowy eksperyment i ponownie testować skutki postawionej hipotezy w świetle nowych danych, itd. Tak więc zgodnie z powyższą regułą, badanie naukowe w zasadzie nigdy się nie kończy, przy czym jeśli pewna hipoteza została sformułowana i sprawdzona, gdy dowiodła swego „hartu", nie wolno jej poniechać bez podania „dobrego powodu", np. na drodze falsyfikacji jednej z konsekwencji hipotezy, por. [116].

Jednak w zastosowaniach technicznych podstawową wadą takiego podejścia jest to, że *nie uwzględnia się celu*, dla którego tworzony jest model. W praktyce bowiem, gdy model jest tworzony dla celów inżynierskich, zamiast kontynuować postępowanie, dopóki wszystkie dane nie zostaną *wyjaśnione* przez model, poszukuje się iteracyjnie modelu dotąd, aż uzyskany zostanie model *odpowiedni do* założonego *celu*. Oba podejścia oparte są na iteracyjnym poszukiwaniu modelu, ale przy różnych warunkach zatrzymania poszukiwań. Na ich podstawie przedstawiona zostanie w punkcie 5.2 ogólna procedura identyfikacji.

2.1.1. Podstawowe wymagania wobec identyfikacji

Na ogół można jawnie wyróżnić dwa przedziały czasu: *etap identyfikacji procesu* oraz *etap stosowania modelu*. Na etapie identyfikacji wykonywany jest *eksperyment identyfikacyjny*, podczas którego obiekt poddawany jest pobudzeniu, jeśli to możliwe, to w sposób specjalnie zaprojektowany dla celów identyfikacji, a zachowanie procesu jest obserwowane. Etap ten kończy się skonstruowaniem modelu matematycznego, który w zwięzły sposób opisuje zachowanie procesu w czasie eksperymentu. Opis taki powinien być *reprezentatywny*, tzn.

model uzyskany na etapie identyfikacji powinien obowiązywać także w czasie jego stosowania i być *odpowiedni do celu* [17]. Warunek ten wiąże się w zasadzie z *niezmiennością* własności procesu. Jeżeli jednak wymaganie to nie może być w pełni spełnione, to dopuszcza się pewną zmienność modelowanej dynamiki procesu, o ile jest ona odpowiednio wolna (w porównaniu ze zmiennością sygnałów) oraz do pewnego stopnia regularna, w tym sensie, że można ją zamodelować.

Opis ten powinien być także odpowiednio *dokładny* ze względu na ten cel. Osiągalna *dokładność* modelu zależy od długości trwania etapu identyfikacji (ściśle biorąc – eksperymentu identyfikacyjnego) oraz wszelkich ograniczeń nałożonych na pobudzenie procesu w trakcie tego eksperymentu. Ograniczenia dokładności wynikające z obu tych czynników można zmniejszyć przy zwiększonym koszcie wykonania eksperymentu (w sensie kosztu pomiarów, produkcji itp.). Dokładność związana jest także z *adekwatnością* przyjętego opisu do własności obiektu, a co za tym idzie z jego reprezentatywnością. Zmieniająca się bowiem charakterystyka procesu może stanowić ograniczenie, którego nie da się przekroczyć. Można przyjąć, że warunek ten jest spełniony w praktyce tylko dla ograniczonego przedziału czasu. A więc jeśli z tego punktu widzenia etap stosowania modelu trwa zbyt długo, to powinien po nim wystąpić nowy etap identyfikacji.

W zależności od ogólnych warunków prowadzenia identyfikacji można wyróżnić identyfikację w trybie *off-line* lub *on-line*. W tym pierwszym przypadku zakłada się, że etapy identyfikacji i stosowania są rozłączne w tym sensie, że najpierw przeprowadza się eksperyment identyfikacyjny, po czym stosując odpowiednią procedurę identyfikacji uzyskuje się model przeznaczony dla jakiegoś celu, a następnie na etapie stosowania model ten jest wykorzystywany do tego celu. Identyfikacja w trybie *off-line* nosi także nazwę identyfikacji *wsadowej* (ang. *batch*), jako że wsad danych jest magazynowany i może być dostępny do wielokrotnego przetwarzania w dowolnym czasie, zanim model zostanie użyty.

W pewnych przypadkach wymagane jest jednak prowadzenie identyfikacji w czasie rzeczywistym, czyli w trybie *on-line*. Dotyczy to sytuacji, gdy przetwarzanie danych jest realizowane na bieżąco, w miarę ich napływania. Wtedy długość etapu identyfikacji (a stąd i dokładność modelu) ograniczona jest jedynie przez warunek *reprezentatywności*. Identyfikacja w trybie *on-line* stosowana jest także powszechnie w metodach samonastrajania regulatorów lub sterowania adaptacyjnego, gdy przedział czasu, dla którego spełniony jest warunek *reprezentatywności*, jest na tyle krótki, że etapy identyfikacji i stosowania zachodzą na siebie. Występują tu więc dwa cele, które są wzajemnie sprzeczne: z jednej strony takie sterowanie obiektem, by wielkość wyjściowa jak najmniej odbiegała od wartości zadanej, a z drugiej takie

pobudzanie obiektu, aby możliwa była jego identyfikacja. Może to prowadzić do powstania wielu problemów w prawidłowym przebiegu identyfikacji, por. np. [85].

Tematyka tej pracy związana jest z identyfikacją prowadzoną w trybie *off-line*. Tu jednak można wyróżnić dwa przypadki. Otóż zachowanie procesu na *etapie stosowania modelu* wynika z jego normalnego przeznaczenia i z uwagi na osiągalną dokładność może nie być odpowiednie dla *etapu identyfikacji*. Stąd normalny przebieg procesu może jednocześnie stanowić pewne ograniczenie na możliwość jego identyfikacji. Jeżeli zadanie da się sformułować w ten sposób, by na etapie identyfikacji proces mógł być pobudzany bardziej efektywnie, w sposób *specjalnie zaprojektowany* na potrzeby eksperymentu, to mamy do czynienia z identyfikacją w układzie *otwartym*. Jest to sytuacja najbardziej pożądana dla celów identyfikacji. Jeżeli jednak w czasie potrzebnym na przeprowadzenie eksperymentu identyfikacyjnego nie można ignorować przeznaczenia procesu, to wtedy sposób jego pobudzania wynika z tego przeznaczenia. Odpowiada to sytuacji, gdy pobudzenie jest generowane w *sposób spontaniczny*, niezależnie od osoby prowadzącej eksperyment, albo gdy obiekt pracuje w układzie *sprzężenia zwrotnego*. To może skutkować otrzymaniem mniej dokładnego modelu lub prowadzić do problemów związanych z tzw. identyfikowalnością systemową, czyli możliwością wyznaczenia współczynników identyfikowanego modelu, por. [43, 62, 79].

2.1.2. Procedura identyfikacji off-line

Bazując na logice eksperymentalnej metody badawczej można przedstawić klasyczny schemat procedury identyfikacji prowadzonej w trybie *off-line*:

1. Wykonanie *eksperymentu* i zebranie pomiarów.
2. *Wybór struktury modelu* rozumianej jako postać, w jakiej model ma być zapisany.
3. *Estymacja* modelu zgodnie z wybranym kryterium *aproksymacji*.
4. *Testowanie* otrzymanego modelu.

Jeżeli wynik identyfikacji nie spełnia wymagań, to należy zmienić strukturę modelu i powtórzyć procedurę od kroku drugiego, aż do uzyskania odpowiedniego modelu.

Jednak na ogół problem jest bardziej złożony, jako że dwa pierwsze etapy są wzajemnie zależne. Struktura modelu powinna być dobrana adekwatnie do wyników eksperymentu, ale z drugiej strony dokładność uzyskanego modelu zależy w dużym stopniu od eksperymentu, który powinien być tak zaplanowany, aby dostarczyć maksimum informacji dla konkretnej struktury. Stąd często procedurę trzeba powtarzać także od pierwszego etapu. Poza tym oba

etapy zależą ściśle od dwóch innych źródeł informacji: wiedzy *a priori* o procesie i *celu*, dla którego tworzony jest model.

Według Poppera teorie nie są weryfikowalne w sensie sprawdzenia ich prawdziwości, ale mogą zostać „potwierdzone" w tym sensie, że można pokazać, jakie *testy falsyfikacji* dana hipoteza przetrwała i w jakiej mierze. Przenosząc to rozumowanie na przypadek tworzenia modelu dla celów inżynierskich, na określenie testów sprawdzających model także, ściśle rzecz biorąc, nie powinno się używać słowa *weryfikacja*, chociaż pojęcie to bywa stosowane w literaturze przedmiotu, por. np. [50, 101]. Testy te bowiem nie pozwalają na potwierdzenie, że model jest *prawdziwy*. Aby podkreślić fakt, że testować można jedynie *odpowiedniość do celu*, w literaturze angielskojęzycznej używane jest określenie *validation*, jako że określenie *odpowiedni do danego celu* jest wymaganiem słabszym niż *prawdziwy*. Także w tej pracy testowanie, czy model jest odpowiedni do wyznaczonego celu, nazywane będzie *walidacją*, co stanowi bezpośrednie tłumaczenie angielskiego *validation*.

Walidacja polega więc na przeprowadzeniu takich testów, których wynik określa, czy model jest *zadowalający ze względu na postawiony cel*. O modelu, który pozytywnie przeszedł testy walidacji, można powiedzieć, że jest *odpowiedni* do postawionego celu (ang. *validated*).

Jednak nawet jeśli istnieje możliwość potwierdzenia, że model jest odpowiedni do danego celu, to korzystanie z podstawowych procedur testowania i poprawiania hipotez na drodze ich falsyfikacji może być również użyteczne. *Falsyfikacja* polega na wykonaniu takich testów, których wynik określa, czy model (lub jego struktura) *nie jest sprzeczny z danymi*. Jeśli model okaże się być sprzeczny z danymi, to powinien być *odrzucony* (*sfalsyfikowany*), por. [17].

Warto tu zauważyć, że model przyjęty jako *odpowiedni do celu* może się okazać prostszy niż ten, który można by uznać za *niesprzeczny* z danymi. Może się więc zdarzyć, że model *odpowiedni do celu*, czyli taki, który przeszedł test walidacji, zostanie jednocześnie *odrzucony*, jeśli nie przeszedł testu falsyfikacji. Wynik taki wcale nie musi stanowić paradoksu, np. do poprawnego opisu jakiegoś procesu przemysłowego może być potrzebny złożony model dynamiczny, odpowiedni do wyjaśnienia danych eksperymentalnych, podczas gdy do celów strojenia regulatora PID zgodnie z regułą Zieglera-Nicholsa może wystarczyć prosty model w postaci inercji pierwszego rzędu z opóźnieniem.

Z drugiej strony tak wyraźny podział testów na dwa rodzaje, jaki zaproponowano w [17], może w praktyce rodzić pewne wątpliwości. O ile bowiem są testy, które można jednoznacznie uznać za testy *falsyfikacji* (klasycznym przykładem może tu być test białości błędów predykcji, por. 5.1.2), to niektóre testy mogą służyć zarówno walidacji, jak i falsyfikacji modelu, w zależności od punktu widzenia. Na przykład, w przypadku testu symulacji, por. 5.4.1, o modelu, dla którego błąd symulacji jest duży, można powiedzieć, że jest *nieodpowiedni do*

celu (bo nie umożliwia symulacji wyjścia z odpowiednią dokładnością, jeśli taki był cel identyfikacji), albo że jest sprzeczny z danymi (bo po prostu dla danych symulacja jest zła). Stąd w podręcznikach zwykle nie rozróżnia się obu grup testów, używając tylko jednego określenia: *model validation*, por. np. [79, 108]. Tak więc, o ile rozróżnienie to jest przydatne z punktu widzenia dyskusji teoretycznych aspektów iteracyjnej procedury identyfikacji przeprowadzonych w punkcie 5.1, to w praktyce może być ominięte i stąd w punkcie 5.4 mówi się po prostu o testowaniu modelu.

2.1.3. Potrzeba komputerowego wspomagania identyfikacji

Przedstawiona w poprzednim punkcie ogólna procedura identyfikacji *off-line* stanowi pewne ramy umożliwiające zaprojektowanie odpowiedniego *oprogramowania wspomagającego* tworzenie modeli. Powszechnie istnieje bowiem zgoda co do tego, że jakaś forma komputerowego wspomagania jest niezbędnym warunkiem pomyślnego przeprowadzenia procedury identyfikacji. Złożoność i szczegóły konkretnego programu zależeć będą od tego, jak ogólne struktury oraz cele program ten będzie obsługiwał, a także jakie metody aproksymacji oraz testowania zostaną w nim zaimplementowane. Oprogramowanie takie powinno oczywiście dostarczać wysokiej jakości procedur numerycznych, niezbędnych do wykonania omówionych faz oraz być wyposażone w odpowiednie narzędzia do przedstawiania wyników w postaci graficznej czy obsługi baz danych.

Jednak aby osiągnąć dobre wyniki, osoba prowadząca identyfikację musi posiadać, oprócz oczywiście informacji *a priori* o identyfikowanym procesie, także jakieś *doświadczenie* w tym względzie i pewną *wiedzę*, które są konieczne szczególnie w fazach *planowania eksperymentu* i *określenia struktury modelu*. Można je uważać za swego rodzaju wiedzę eksperta, która jest wprawdzie trudna do precyzyjnego zdefiniowania, niemniej jednak bardzo realna. Stąd oczywiste jest oczekiwanie, aby oprogramowanie do komputerowego wspomagania identyfikacji zawierało nie tylko standardowe algorytmy przetwarzania czy obsługi baz danych, ale szło dalej, tzn. dostarczało pewnych usług typowych dla systemów *eksperckich*, por. [99].

Przyglądając się bliżej zadaniom koniecznym do wykonania w kolejnych etapach procedury identyfikacji oraz uwzględniając fakt, że na ogół dane z eksperymentu identyfikacyjnego powinny zostać przygotowane do dalszego przetwarzania, można stwierdzić, że na wiedzę, doświadczenie i umiejętności osoby wykonującej identyfikację składają się w szczególności:

- umiejętność planowania eksperymentu identyfikacyjnego;
- umiejętność odpowiedniego przygotowania danych,

- umiejętność określenia struktury modelu,
- umiejętność wyboru odpowiedniej procedury numerycznej dla aproksymacji modelu oraz parametryzacji tej procedury,
- umiejętność interpretowania wyników obliczeń numerycznych,
- umiejętność testowania i analizowania wyników identyfikacji,
- umiejętność łączenia szeregu różnych procedur numerycznych i graficznych w jedno użyteczne i efektywne narzędzie.

Wiele z wymienionych wyżej umiejętności jest słabo zalgorytmizowanych, jednak niektóre z nich mogą być wyrażone w sposób bardziej formalny, za pomocą reguł heurystycznych opartych na odpowiednio bogatym doświadczeniu symulacyjnym i eksperymentatorskim. Oznacza to, że mogą być one zaprogramowane w komputerze. Stąd w dalszej części przyjmuje się w domyśle, że jeżeli jest mowa o programie do komputerowego wspomagania identyfikacji, to jednocześnie zawarte są w nim pewne usługi, które można by nazwać *inteligentnymi*. Oczywiście, nie oznacza to, że program taki może w całości zastąpić człowieka w procedurze identyfikacji, ale że inteligentnie wspomaga on użytkownika w realizacji tej procedury w takim zakresie, w jakim to jest możliwe.

2.2. Identyfikacja w ujęciu probabilistycznym

Nawet w przypadku wystąpienia takiego samego pobudzenia obserwowane odpowiedzi rzeczywistych procesów zazwyczaj różnią się między sobą. Jeżeli różnice w odpowiedzi procesu nie są spowodowane jego niestacjonarnością, którą można modelować, to wtedy można przyjąć, że odpowiedź taka jest *losowa*. Z drugiej strony, w odpowiedzi zawarta jest także pewna informacja, która jest niezmienna i którą można wykorzystać. Tak więc zadaniem identyfikacji, podobnie jak całego wnioskowania statystycznego, jest wydobycie tej informacji z obserwacji, które same w sobie mogą być niepewne. Problem *modelowania* polega na przedstawieniu tej informacji w postaci przydatnej do wykorzystania, np. jako struktura czy parametry modelu, przy wykorzystaniu pojęć z dziedziny rachunku prawdopodobieństwa i statystyki, takich jak np. zmienna losowa, wzór Bayesa czy funkcja wiarygodności.

2.2.1. Wzór Bayesa

Niech zmienna losowa $\mathcal{X}$ jest odwzorowaniem z danej przestrzeni zdarzeń elementarnych Ω do przestrzeni możliwych wartości x, tj. $x = \mathcal{X}(\varpi)$, gdzie $\varpi \in \Omega$. Przestrzeń Ω ma dołączoną miarę

prawdopodobieństwa lub rozkład prawdopodobieństwa dla ϖ, które determinują rozkład zmiennej losowej $\mathcal{X}$. Każde *zdarzenie* generuje konkretną wartość ϖ (z danym prawdopodobieństwem), a wartość $\mathbf{x} = \mathcal{X}(\varpi)$ jest wynikiem zdarzenia, *realizacją*.

Gęstość prawdopodobieństwa $p(x)$ zmiennej losowej $\mathcal{X}$ można uważać za formę informacji, jaka jest dostępna na temat $\mathcal{X}$ - im rozkład bardziej skupiony, tym bardziej dokładna informacja o $\mathcal{X}$. Jeżeli brak jest możliwości obserwacji zdarzenia $\mathbf{x}$, a gęstość prawdopodobieństwa $p(x)$ nie jest dokładnie znana, natomiast można przeprowadzić *doświadczenie* polegające na obserwacji innej zmiennej losowej $\mathcal{Y}$, która jest zależna od $\mathcal{X}$, to o rozkładzie zmiennej losowej $\mathcal{X}$ można będzie powiedzieć coś więcej, niż założona *a priori* gęstość $p(x)$. Po wykonaniu doświadczenia (obserwacji $\mathbf{y}$ zmiennej losowej $\mathcal{Y}$) informacja na temat $\mathcal{X}$ stanowi rozkład *a posteriori* $p(x \mid \mathbf{y})$, przy czym jeśli $\mathbf{y}$ jest rzeczywiście zależne od $\mathbf{x}$, to rozkład warunkowy $p(x \mid \mathbf{y})$ jest bardziej skupiony niż $p(x)$, co wynika ze wzoru Bayesa [136]:

$$p(x \mid \mathbf{y}) = p(\mathbf{y} \mid x)\, p(x) / p(\mathbf{y}). \tag{2.1}$$

Przedstawione wyżej rozważania można przenieść na grunt modelowania i identyfikacji przyjmując, że odpowiednikiem zmiennej losowej $\mathcal{X}$ jest identyfikowany proces, natomiast generowane przez ten proces sygnały to obserwowana zmienna $\mathcal{Y}$. Jeżeli dla procesu można wyróżnić zmienne *wejściowe* (inaczej *pobudzenie* lub zmienne *sterujące*) oraz zmienne *wyjściowe*, będące odpowiedzią na zmienne wejściowe, to nazywa się go *obiektem*. Zakłada się jednocześnie, iż to, co jest wejściem, a co wyjściem, jest dane *a priori*. Odpowiedź obiektu może zależeć również od innych, nieznanych zmiennych (traktowanych jako zakłócenia). Ponieważ wejścia oraz wyjścia mogą być zarówno skalarami, jak i wektorami, stąd wyróżnia się obiekty *SISO* (*Single-Input-Single-Output*), *MISO* (*Multi-Input-Single-Output*) oraz *MIMO* (*Multi-Input-Multi-Output*).

W celu dalszej dyskusji zmienną wyjściową oznaczać się będzie przez y, jej obserwację przez $\mathbf{y}$, natomiast sam proces jako $\mathbf{P}$. Odpowiednio wejście oznaczane będzie przez u, natomiast proces generujący konkretne pobudzenie $\mathbf{u}$ jako $\mathbf{R}$. Tak więc źródłem danych $(\mathbf{u}, \mathbf{y})$ jest para $(\mathbf{R}, \mathbf{P})$, którą można uważać za system doświadczalny.

Z przyjętych definicji wynika efekt przyczynowo-skutkowy: $(\mathbf{R}, \mathbf{P}) \rightarrow (\mathbf{u}, \mathbf{y})$. Ponieważ na ogół celem identyfikacji jest modelowanie obiektu generującego $\mathbf{y}$ w odpowiedzi na pobudzenie $\mathbf{u}$, *niezależnie* od źródła tego pobudzenia, stąd bardzo często w teorii identyfikacji zakłada się (jawnie lub niejawnie), że źródło pobudzenia i obiekt mogą być *rozdzielone*, tzn. z relacji $(\mathbf{R}, \mathbf{P}) \rightarrow (\mathbf{u}, \mathbf{y})$ wynika, że $\mathbf{R} \rightarrow \mathbf{u}$ oraz $(\mathbf{u}, \mathbf{P}) \rightarrow \mathbf{y}$, por. [17]. Jest to o tyle istotne, że na etapie

stosowania modelu źródło pobudzenia może być inne niż na etapie identyfikacji. Jeśli w rzeczywistości założenie to nie będzie spełnione, to wtedy wynik identyfikacji będzie błędny, por. 5.2.2.

Traktując system doświadczalny (**R**, **P**) jako realizację zmiennej losowej (R, P), podobnie jak przez **x** oznaczono realizację zmiennej losowej x, oraz zbiór danych (**u**, **y**) jako dostępną obserwację systemu (**R**, **P**), można sformułować zadanie identyfikacji korzystając ze wzoru Bayesa

$$p(R, P \mid \mathbf{u}, \mathbf{y}) = p(\mathbf{u}, \mathbf{y} \mid R, P)\, p(R, P) / p(\mathbf{u}, \mathbf{y}). \tag{2.2}$$

W *podejściu bayesowskim* uważa się więc zarówno dane, jak i system doświadczalny za zmienne losowe, których zależność jest określona przez ich łączną gęstość prawdopodobieństwa, por. [114].

Jeżeli proces generujący pobudzenie **R** jest znany, to wykorzystując wzór na gęstość prawdopodobieństwa warunkowego $p(P \mid R) = p(R, P) / p(R)$ oraz fakt, że prawdopodobieństwo tego, że $R = \mathbf{R}$ wynosi 1, można (2.2) przekształcić do postaci

$$p(P \mid \mathbf{R}, \mathbf{u}, \mathbf{y}) = p(\mathbf{u}, \mathbf{y} \mid \mathbf{R}, P)\, p(P \mid \mathbf{R}) / p(\mathbf{u}, \mathbf{y} \mid \mathbf{R}). \tag{2.3}$$

Jest to przypadek szczególnie pożądany, jako że zasadniczym celem identyfikacji jest znalezienie modelu procesu. Ponieważ jednak nie zawsze **R** jest znane w pełni, stąd aby móc uzyskać model samego procesu, musi być spełniony warunek rozdzielności, co będzie omówione bliżej dla przypadku identyfikacji obiektu dynamicznego, por. 2.2.4.

Ocenę procesu **P** wyznacza się więc na drodze maksymalizacji gęstości prawdopodobieństwa *a posteriori* $p(P \mid \mathbf{R}, \mathbf{u}, \mathbf{y})$. Jest to tzw. ocena *MAP* (*Maximum a Posteriori*) [79]. Do jej wyznaczenia wykorzystuje się wartość funkcji gęstości prawdopodobieństwa $p(\mathbf{u}, \mathbf{y} \mid \mathbf{R}, P)$, tj. wartość, jaką miałyby dane $(\mathbf{u}, y)$, gdyby pochodziły z systemu $(\mathbf{R}, P)$, obliczoną dla zaobserwowanych danych (**u**, **y**) oraz to, co *a priori* wiadomo o rzeczywistym procesie, czyli $p(P \mid \mathbf{R})$. Natomiast skoro P nie występuje jako argument w $p(\mathbf{u}, \mathbf{y} \mid \mathbf{R})$, więc z punktu widzenia estymacji **P** czynnik ten nie ma znaczenia i pełni jedynie rolę współczynnika normalizującego [17].

2.2.2. Funkcja wiarygodności

Do rozwiązania zadania identyfikacji polegającego na wydobyciu jak najwięcej informacji z uzyskanych danych można także wykorzystać wprowadzoną przez Fishera *funkcję wiarygodności* [33]. W tym przypadku tylko obserwacje (**u**, **y**) uważa się za realizacje zmiennych losowych, a model procesu wybiera się tak, aby zajście zaobserwowanego zdarzenia

było jak „najbardziej prawdopodobne". Stąd gęstość prawdopodobieństwa $p(\mathbf{u}, \mathbf{y} \mid \mathbf{R}, P)$ traktuje się jako „wiarygodność" tego, że zdarzenie to rzeczywiście zaszło i w praktyce wyznaczenie modelu dla **P** odbywa się na drodze maksymalizacji funkcji wiarygodności:

$$L(P \mid \mathbf{R}, \mathbf{u}, \mathbf{y}) = p(\mathbf{u}, \mathbf{y} \mid \mathbf{R}, P). \tag{2.4}$$

Uzyskana w ten sposób ocena nosi nazwę oceny *ML* (*Maximum Likelihood*) [79]. Różnica w stosunku do podejścia bayesowskiego polega na tym, że pomija się gęstość prawdopodobieństwa *a priori* $p(P \mid \mathbf{R})$, którą można także traktować jako prawdopodobieństwo subiektywne, por. [120]. W szczególności, jeśli o P wiadomo bardzo mało, to można założyć, że w całej podanej przestrzeni, w której P może występować, wszystkie P są tak samo prawdopodobne, a więc mają rozkład lokalnie jednostajny (w przybliżeniu jednostajny w przestrzeni, w której funkcja wiarygodności jest istotna) [19]. Poza tym, dla średniej lub dużej liczby danych (rzędu setki czy tysiące) wpływ prawdopodobieństwa *a priori* na obliczone prawdopodobieństwo *a posteriori* jest niewielki [25]. Stąd funkcja wiarygodności i funkcja gęstości rozkładu *a posteriori* są proporcjonalne i to, co uważa się za *wiarygodność* procesu, jest gęstością prawdopodobieństwa *a posteriori* danych obserwowanych (**u**, **y**), generowanych przez pewien system (**R**, P).

Ponieważ w identyfikacji opartej bezpośrednio na wzorze Bayesa wyznacza się model na drodze maksymalizacji gęstości prawdopodobieństwa *a posteriori* uwzględniając oba czynniki $p(\mathbf{u}, \mathbf{y} \mid \mathbf{R}, P)$ oraz $p(P \mid \mathbf{R})$, więc metodę *MAP* można uważać za uogólnienie metody *ML* [70, 69]. *MAP* wymaga jednak przedstawienia wiedzy *a priori* o (R, P) w takiej postaci, aby można było przyjąć jakąś rozsądną interpretację dla gęstości $p(P \mid \mathbf{R})$. W identyfikacji opartej na maksymalizacji funkcji wiarygodności (2.4) problem ten się omija, stąd znacznie większa popularność metody *ML*.

Warto tu jednak zauważyć, że rozkład prawdopodobieństwa *a posteriori* stanowi jedynie opis eksperymentu, tzn. model uzyskany na tej podstawie będzie odpowiedni tylko dla tego konkretnego eksperymentu. Tymczasem oczekuje się, że model ten będzie odpowiedni także na etapie stosowania, a przynajmniej dla pobudzenia u o takich własnościach, które będą dominowały na tym etapie. Wymaganie to może być trudne do spełnienia, jeśli eksperyment identyfikacyjny przeprowadzony był w sposób nieprawidłowy, a nawet może się okazać nieosiągalne. Taką przeszkodą jest np. brak odpowiednio *dużej liczby* doświadczeń, aby można było stwierdzić, że model jest *niezależny* od eksperymentu. W szczególności dla procesów niestacjonarnych, dla których odpowiedź zależy od upływającego czasu, może w ogóle wystąpić brak możliwości powtórzenia doświadczenia, jak np. w przypadku analizy danych historycznych dla celów modelowania procesów ekonomicznych czy socjologicznych. Stąd

w rozdziale 3 przedstawiono bliżej wymagania wobec eksperymentu identyfikacyjnego, aby wynikające z funkcji wiarygodności lub wzoru Bayesa wnioskowanie statystyczne pozwoliło na uzyskanie opisu adekwatnego także na etapie stosowania.

W zadaniu identyfikacji trzeba przyjąć pewne *założenia* odnośnie do modelowanego procesu i źródła pobudzenia, co na ogół odbywa się na drodze zdefiniowania odpowiedniego *modelu probabilistycznego*, np. funkcji $p(y \mid u)$ określającej gęstość prawdopodobieństwa dla zmiennej y uwarunkowanej zmienną u. Są tu dwie możliwości przedstawienia takich założeń. Po pierwsze założenia o *zależności* lub *niezależności* poszczególnych zmiennych można wyrazić przez wprowadzenie ich lub usunięcie z listy po prawej stronie znaku warunku |, np. jeśli *a priori* zakłada się, że zmienna z nie wpływa na y, to przyjmuje się, że $p(y \mid u, z) = p(y \mid u)$. Druga możliwość to przyjęcie *a priori* jakiejś *struktury* modelu przez przedstawienie $p(y \mid u)$ w odpowiedniej postaci, co zostanie omówione bliżej w 4.1.1.

2.2.3. Eksperyment identyfikacyjny

Eksperyment identyfikacyjny polega na poddaniu procesu **P** pobudzeniu **u** i obserwacji jego odpowiedzi **y** w pewnym okresie czasu **T**. Można więc powiedzieć, że eksperyment wykonywany jest przez „eksperymentatora" **R**, który *generuje pobudzenie* **u** (pełni rolę źródła pobudzenia z punktu 2.2.1) i *rejestruje* odpowiedź **y**. Zazwyczaj w przypadku identyfikacji procesów o charakterze technicznym jest to komputer sprzęgnięty poprzez układ wejść-wyjść analogowych z elementami wykonawczymi oraz czujnikami pomiarowymi. Rolę tę może oczywiście pełnić jakiekolwiek inne urządzenie lub człowiek (np. operator procesu), o ile jest w stanie wykonać oba zadania, tj. pobudzanie i rejestrację. Pobudzenie **u** może stanowić zbiór sygnałów testujących (deterministycznych) generowanych przez **R** i wtedy **R** jest całkowicie znane. Jednak jeżeli **u** jest generowane w sposób spontaniczny, nieznany dokładnie, to **R** może być znane tylko częściowo. Pobudzenie **u** może także w ogóle nie występować i wtedy obserwowany jest tylko „naturalny" proces.

Bardzo ważne jest tu rozróżnienie, co jest uważane za **R**, a co za **P**. Ich wzajemne oddziaływanie odbywa się poprzez **u** i **y**, które reprezentują dane, a nie zmienne fizyczne. Tak więc do procesu powinno się zaliczać także wszystko to, co się znajduje w torze przesyłania i przekształcania sygnałów, a w szczególności elementy wykonawcze oraz czujniki pomiarowe. Podobnie układ wejść-wyjść analogowych powinien być w zasadzie traktowany jako element procesu, a nie eksperymentatora.

Ponieważ odpowiedź **y** jest nie tylko efektem pobudzenia **u**, więc oprócz wymienionych tu **R** i **P**, w eksperymencie identyfikacyjnym można wyróżnić także otoczenie **O**. Stanowi ono resztę otaczającego świata, która może wpływać na **P**, ale także na **R**. Z punktu widzenia **R** trudno jest rozróżnić, na ile odpowiedź **y** jest wynikiem pobudzenia **u**, a na ile wpływu otoczenia **O**. Tak więc proces i otoczenie mogą być traktowane wspólnie jako system $\mathbf{S} = (\mathbf{O}, \mathbf{P})$ będący przedmiotem identyfikacji, jak np. w [79] czy [132], aczkolwiek **S** można także interpretować jako „świat otaczający" **R**, jak np. w [17]. W dalszym ciągu zamiast o procesie **P** będzie więc mowa o *obiekcie identyfikacji* **S** rozumianym w przedstawionym wyżej sensie.

Dla danych doświadczalnych $(\mathbf{u}, \mathbf{y})$ można wyznaczyć warunkowy rozkład prawdopodobieństwa $p(y \mid \mathbf{u}, \mathbf{R}, \mathbf{S}, \mathbf{T})$ i jest to maksymalna ilość informacji, jaką można uzyskać na podstawie eksperymentu. Potrzebny jest natomiast rozkład dla y występujących na etapie *stosowania*, tj. w czasie T. Przyjmując założenie, że zarówno proces, jak i otoczenie pozostaną niezmienione, to rozkład ten można przedstawić jako $p(y \mid u, R, \mathbf{S}, T)$, gdzie R oznacza proces sterujący na etapie stosowania. Na ogół brak jest jednak pewności, że wyznaczona na podstawie eksperymentu funkcja $p(y \mid \mathbf{u}, \mathbf{R}, \mathbf{S}, \mathbf{T})$ mówi coś na temat rozkładu $p(u, y \mid R, \mathbf{S}, T)$ dla etapu stosowania. Rozkład ten opisuje równanie $p(u, y \mid R, \mathbf{S}, T) = p(u \mid R, \mathbf{S}, T)\, p(y \mid u, R, \mathbf{S}, T)$ i nawet jeśli sterowanie będzie niezależne od **S**, to i tak odpowiedź będzie scharakteryzowana przez $p(y \mid u, R, \mathbf{S}, T)$, a nie przez $p(y \mid \mathbf{u}, \mathbf{R}, \mathbf{S}, \mathbf{T})$. Ponieważ po znaku warunku | występują trzy wielkości, które na etapie stosowania mogą być inne, tj. (u, R, T), więc jeśli model ma być użyteczny na tym etapie, to trzeba być pewnym, że jest on adekwatny także *poza* zakresem zmiennych $(\mathbf{u}, \mathbf{R}, \mathbf{T})$ występujących w czasie eksperymentu identyfikacyjnego, por. [17]. Podstawowe wymagania wobec eksperymentu identyfikacyjnego, które pozwolą na przeprowadzenie prawidłowego wnioskowania dla obiektu dynamicznego, omówiono w punkcie 3.1.

2.2.4. Obiekt dynamiczny

Dla obiektów dynamicznych pobudzenie i odpowiedź stanowią *ciągi* danych

$$U_N = \{u(i) \mid i = 1, 2, ..., N\}, \qquad Y_N = \{y(i) \mid i = 1, 2, ..., N\}, \tag{2.5}$$

które nazywa się też ciągami czasowymi (odpowiednio wejściowym i wyjściowym) [101], jako że są one funkcjami czasu dyskretnego i. Ciągi te mogą być skalarne lub wektorowe, co stanowi podstawę klasyfikacji wymiarowości obiektów, por. 2.2.1.

Oznaczając przez U_{i-1} ciąg pobudzeń $u(1), u(2), \ldots, u(i-1)$, a przez Y_i ciąg odpowiedzi $y(1), y(2), \ldots, y(i)$, gęstość prawdopodobieństwa dla odpowiedzi w chwili i-tej można przedstawić jako zależność warunkową od ciągu poprzednich pobudzeń oraz odpowiedzi $p[y(i) \mid U_{i-1}, Y_{i-1}, R, S]$. Podobnie jest dla pobudzenia, przy czym jeżeli $p[u(i) \mid U_{i-1}, Y_i, R, S]$ rzeczywiście zależy od Y_i, to taki eksperyment jest prowadzony w *układzie zamkniętym*.

W powyższych wyrażeniach przyjęto naturalny dla komputerowych układów sterowania sposób indeksacji próbek, tzn. że sterowanie w chwili bieżącej $u(i)$ wyznacza się na podstawie zmierzonego $y(i)$, które przypisuje się do tej samej chwili i.

Gęstość prawdopodobieństwa dla całej próbki danych (U_N, Y_N) wynosi $p(U_N, Y_N \mid R, S)$, a po wielokrotnym zastosowaniu wzoru Bayesa jest:

$$p(U_N, Y_N \mid R, S) = p[u(N), y(N) \mid U_{N-1}, Y_{N-1}, N, R, P]\, p[U_{N-1}, Y_{N-1} \mid R, S] = \\ = \prod_{i=1}^{N} p[u(i), y(i) \mid U_{i-1}, Y_{i-1}, R, S]. \tag{2.6}$$

Na podstawie eksperymentu można uzyskać model systemu obejmującego *łącznie* obiekt i źródło pobudzenia [17]. Ponieważ na ogół celem identyfikacji jest uzyskanie modelu obiektu, stąd pojawia się postulat ich rozdzielenia.

Aby można było to zrobić, powinna być spełniona zależność $p(u, y \mid \mathbf{R}, \mathbf{S}, \mathbf{T}) = p(u \mid \mathbf{R}, \mathbf{T})\, p(y \mid u, \mathbf{S}, \mathbf{T})$, a więc w trakcie trwania eksperymentu wyjście y i pobudzenie u obiektu nie mogą zależeć od jakiejś wspólnej (ukrytej) zmiennej. Najprostszym sposobem spełnienia tego wymagania jest wygenerowanie pobudzenia w sposób całkowicie niezależny od **S**. Można to uzyskać wtedy, gdy U_N jest z góry ustalonym ciągiem, niezależnym od wszystkiego, co może się zdarzyć w czasie eksperymentu, albo gdy pobudzenie jest generowane przez urządzenie, które jest odseparowane od obiektu. Warunek ten spełnia również sytuacja, gdy pobudzenie jest ustalane w czasie eksperymentu (np. przez operatora), ale bez uwzględnienia odpowiedzi obiektu.

Jednak dla obiektów dynamicznych rozdzielenie **R** i **S** jest możliwe także w przypadku wykonywania eksperymentu w układzie zamkniętym, por. [17], albowiem wystarczy, aby **R** nie wpływał na odpowiedź $y(i)$ inaczej jak tylko poprzez poprzedni *ciąg pobudzeń* U_{i-1}, a **S** nie wpływał na pobudzenie $u(i)$ inaczej jak tylko przez *ciąg odpowiedzi* Y_i do chwili bieżącej. Wtedy bowiem dla wszystkich U_i oraz Y_i zachodzi

$$p[u(i) \mid U_{i-1}, Y_i, \mathbf{R}, \mathbf{S}] = p[u(i) \mid U_{i-1}, Y_i, \mathbf{R}],$$

$$p[y(i) \mid U_{i-1}, Y_{i-1}, \mathbf{R}, \mathbf{S}] = p[y(i) \mid U_{i-1}, Y_{i-1}, \mathbf{S}]. \tag{2.7}$$

Warunkiem wystarczającym jest tu więc niezależność sterowania od obiektu, z wyjątkiem przeszłych danych. Jednak w tym przypadku występują jeszcze dodatkowe wymagania na eksperyment, związane z możliwością pojawienia się osobliwości w zadaniu aproksymacji, por. [43, 62, 79].

Jeżeli natomiast obiekt jest pobudzany w sposób naturalny, to warunek rozdzielenia sprowadza się do tego, by źródło pobudzenia nie wpływało na odpowiedź y w inny sposób jak tylko poprzez wejście u. W praktyce jednak może to stwarzać pewne trudności, ponieważ

zwykle nie są znane wszystkie możliwe źródła zmiany pobudzenia i może się okazać, że istnieje zmienna ukryta, wpływająca jednocześnie i na pobudzenie u, i na wyjście y.

Z przedstawionego wyżej warunku rozdzielności **R** i **S** wynika, że

$$p[u(i), y(i) \mid U_{i-1}, Y_{i-1}, \mathbf{R}, \mathbf{S}] = p[u(i) \mid U_{i-1}, Y_i, \mathbf{R}]\, p[y(i) \mid U_{i-1}, Y_{i-1}, \mathbf{S}], \tag{2.8}$$

co oznacza, że jeśli można wyznaczyć $p[y(i) \mid U_{i-1}, Y_{i-1}, \mathbf{S}]$, które powinno być takie samo (teoretycznie) dla każdego eksperymentu, to wynik ten będzie *niezależny* od eksperymentu. Co więcej, na etapie *stosowania* zależność (2.8) przyjmie postać $p[u(i), y(i) \mid U_{i-1}, Y_{i-1}, R, \mathbf{S}] = p[u(i) \mid U_{i-1}, Y_i, R]\, p[y(i) \mid U_{i-1}, Y_{i-1}, \mathbf{S}]$, a więc mimo że obiekt będzie pobudzany inaczej niż w trakcie eksperymentu identyfikacyjnego, to cały system nadal będzie modelowany prawidłowo.

Warto tu także zauważyć, że w modelach probabilistycznych występujących w (2.8) zawarta jest już pewna informacja *a priori* o przyczynowości. Zamiast np. $p[y(i) \mid U_{i-1}, Y_{i-1}, \mathbf{S}]$ można sobie także wyobrazić zapis typu $p[y(i) \mid U_k, Y_{i-1}, \mathbf{S}]$, gdzie $k > i$. Jednak trudno by było przyjmować zależność $y(i)$ od przyszłych wartości pobudzenia $u(i+1), \ldots, u(k)$, chociaż w przypadku istnienia *sprzężenia zwrotnego* tego typu *statystyczna zależność* może faktycznie się pojawić [17].

W przypadku identyfikacji tylko obiektu **S**, jeżeli spełniony jest omawiany tu warunek rozdzielności obiektu i źródła pobudzenia, funkcję wiarygodności dla obiektu dynamicznego można wyrazić wzorem

$$L(S \mid \mathbf{U}_N, \mathbf{Y}_N) = \prod_{i=1}^{N} p[\mathbf{y}(i) \mid \mathbf{U}_{i-1}, \mathbf{Y}_{i-1}, S]. \tag{2.9}$$

Tak więc, bez względu na sposób pobudzania **R**, funkcję wiarygodności maksymalizuje się tak, jakby pobudzenie było z góry zadanym ciągiem *deterministycznym*. Jeżeli jednak wspomniany warunek rozdzielności nie byłby spełniony, to interpretacja wyniku tej optymalizacji nie będzie prawidłowa.

Maksymalizacja funkcji wiarygodności (2.9) może jednak stwarzać problemy numeryczne dla bardzo dużych N, stąd w praktyce większość metod identyfikacji dla systemów dynamicznych bazuje na ocenie *ML* polegającej na minimalizacji funkcji

$$V(S \mid \mathbf{U}_N, \mathbf{Y}_N) \equiv -\ln L(S \mid \mathbf{U}_N, \mathbf{Y}_N)/N. \tag{2.10}$$

Minimalizacja funkcji V stanowić będzie kryterium w zadaniu aproksymacji, por. 5.1.1, więc można ją uważać za *funkcję strat*. Jednak w celu jej minimalizacji najpierw trzeba określić *przestrzeń*, w której S może występować, co jest równoważne z określeniem *struktury modelu*, por. rozdział 4.

Ocena *ML* uzyskana za pomocą minimalizacji funkcji strat (2.10) jest związana z oceną *MAP* wynikającą z podejścia bayesowskiego. Jeżeli bowiem gęstość prawdopodobieństwa *a priori* będzie tak wybrana, że jej wpływ będzie niewielki, to ocena *ML* będzie w przybliżeniu równa ocenie *MAP* [79].

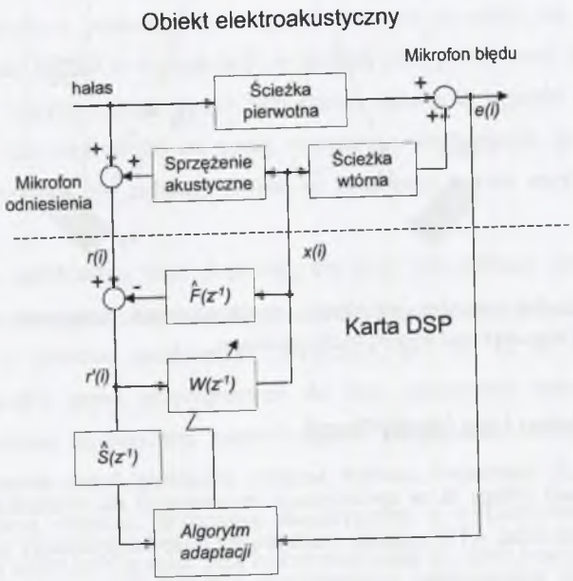
2.3. Przykład obiektu elektroakustycznego

W celu ilustracji przedstawionych rozważań teoretycznych na przykładzie praktycznym rozpatrywane będzie zadanie identyfikacji obiektu elektroakustycznego dla celów aktywnego tłumienia hałasu (ATH). ATH polega na tłumieniu niepożądanych dźwięków za pomocą urządzeń elektroakustycznych przy wykorzystaniu zjawiska interferencji [84]. W literaturze omawianych jest wiele różnych metod ATH, por. np. [47, 72, 95]. We współczesnych zastosowaniach ATH, których celem jest tworzenie lokalnych stref ciszy w pomieszczeniach, wykorzystywane są głównie adaptacyjne cyfrowe układy kompensacji. Prawidłowa parametryzacja takiego układu wymaga znajomości odpowiednich modeli obiektów elektroakustycznych. Jednocześnie własności dynamiczne przestrzeni akustycznej w pomieszczeniu z pogłosem są bardzo złożone i trudne do opisania za pomocą modeli fizycznych, stąd praktycznie *jedynym* sposobem na ich uzyskanie pozostaje wykorzystanie możliwości, jakie daje *identyfikacja.*

2.3.1. Układ kompensacji ATH

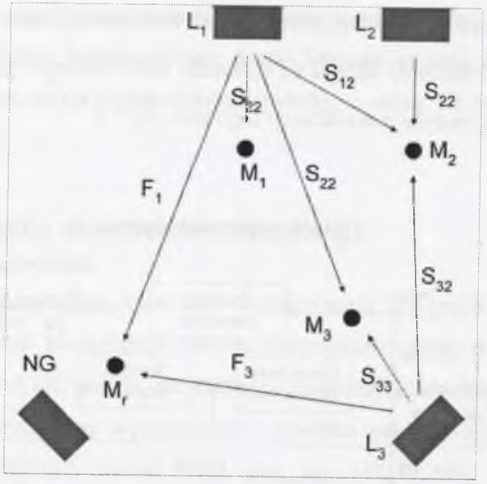
Na rysunku 2.1 przedstawiono schemat blokowy typowego *jednowymiarowego* układu kompensacji dla tworzenia strefy ciszy w pomieszczeniu, por. [72].

Zadanie kompensacji realizowane jest tu z wykorzystaniem karty *DSP* (*Digital Signal Processing*), w której zaimplementowano algorytm adaptacji oraz odpowiednie filtry cyfrowe. Zakłada się, że układ ten pracuje z okresem próbkowania τ_p, i w dalszej części sygnały (podobnie jak i modele) są traktowane jako dyskretne w czasie. Do pomieszczenia dociera hałas, który powinien być skompensowany za pomocą fali akustycznej generowanej przez głośnik sterowany za pomocą sygnału $x(i)$. Celem kompensacji jest uzyskanie maksymalnego tłumienia dźwięku wokół *mikrofonu błędu*, którego sygnał wyjściowy $e(i)$ oraz pochodzący z *mikrofonu odniesienia* sygnał $r(i)$ wykorzystuje się do adaptacji parametrów *filtru kompensacyjnego* $W(z^{-1})$. *Ścieżka pierwotna* oznacza tu przestrzeń akustyczną pomiędzy oboma mikrofonami, natomiast *ścieżka wtórna* reprezentuje wpływ sygnału $x(i)$ na sygnał $e(i)$. Ponieważ dźwięk z głośnika dociera także do mikrofonu odniesienia, stąd powstałe *sprzężenie akustyczne* powinno być w układzie skompensowane. W tym celu od sygnału odniesienia $r(i)$ odejmuje się sygnał $x(i)$ filtrowany przez model transmitancji toru sprzężenia $\hat{F}(z^{-1})$. Aby uzyskać niezależność działania układu od dynamiki toru wtórnego, sygnał $r'(i)$ jest dodatkowo filtrowany za pomocą modelu transmitancji tego toru $\hat{S}(z^{-1})$.



Rys. 2.1. Schemat blokowy układu kompensacji ATH
Fig. 2.1. Block diagram of a feedforward ANC system

W [15] pokazano, że zwiększenie wymiarów strefy ciszy lub tworzenie kilku takich stref możliwe jest przy zwiększeniu liczby układów kompensacji. W dalszym ciągu rozpatrywane będzie zadanie identyfikacji konieczne do sparametryzowania wielowymiarowego układu ATH składającego się z trzech pętli kompensacji, por. [90]. W ten sposób 3 głośniki sterujące oraz 3 mikrofony błędu i 1 mikrofon odniesienia tworzą razem 9 torów wtórnych S_{kl} $(k, l = 1, 2, 3)$ i 3 tory sprzężenia akustycznego F_k, gdzie k oznacza numer głośnika, a l numer mikrofonu błędu. Na rysunku 2.2 pokazano schemat rozmieszczenia elementów sterujących (głośniki L_1 do L_3) i pomiarowych (mikrofony błędu M_1 do M_3 oraz mikrofon odniesienia M_r) w pomieszczeniu. Przez NG oznaczono głośnik generujący zakłócenie (hałas, który ma być tłumiony). Aby nie zaciemniać rysunku, zaznaczono tylko niektóre z torów wtórnych S_{kl} i torów sprzężenia F_k.



Rys. 2.2. Schemat ideowy rozmieszczenia elementów wykonawczych i pomiarowych w pomieszczeniu
Fig. 2.2. Lay-out of actuators and sensors in the enclosure

2.3.2. Opis obiektu i cel identyfikacji

Zadanie identyfikacji polega tu na aproksymacji transmitancji dla wszystkich wymienionych wyżej torów, zanim układ ATH zostanie uruchomiony (etap identyfikacji), po czym uzyskane modele posłużą do odpowiedniej parametryzacji układów kompensacji (etap stosowania). Identyfikacja wykonana będzie więc w trybie *off-line*, a obiektem **S** jest 12 torów elektroakustycznych. Z punktu widzenia algorytmu sterowania istotne jest modelowanie zależności między wygenerowanymi w karcie DSP sygnałami sterującymi a dostarczonymi do karty sygnałami pomiarowymi z mikrofonów, stąd wszystkie urządzenia występujące w układzie ATH (oraz przestrzeń w pomieszczeniu) traktowane są jako elementy obiektu **S**. Każdy tor elektroakustyczny składa się tu z przetwornika cyfrowo-analogowego, dolnoprzepustowego filtru rekonstruującego, wzmacniacza akustycznego, głośnika sterującego, przestrzeni akustycznej między głośnikiem a odpowiednim mikrofonem, mikrofonu, filtru antystroboskopowego i przetwornika analogowo-cyfrowego. Wejściem obiektu u jest więc wektor zawierający 3 sygnały pobudzające dla poszczególnych torów (na rys. 2.1 sygnał taki oznaczono przez x), a wyjściem y wektor zawierający 4 przefiltrowane i spróbkowane sygnały z mikrofonów – trzech mikrofonów błędu (na rys. 2.1 sygnał taki oznaczono przez e) oraz jednego mikrofonu odniesienia (sygnał r). Rolę eksperymentatora **R** pełni tu karta DSP, która później będzie wykorzystywana do kompensacji hałasu.

Omawiany system ATH jest przeznaczony do tłumienia szumów niskoczęstotliwościowych w zakresie około 40 – 140 Hz przy częstotliwości próbkowania f_p = 500 Hz (τ_p = 0.002 s). Filtry

analogowe zaprojektowano jako filtry Butterwortha ósmego rzędu o częstotliwości odcięcia (-3 dB) równej 150 Hz (co odpowiada tłumieniu około 34 dB dla częstotliwości Nyquista). Oba filtry potrzebne są ze względu na zjawiska pojawiające się w układach akustycznych z cyfrowym sterowaniem, por. [113]. Sterowanie zapewnia karta DSP DS1104 R&D firmy dSPACE [90].

Ponieważ sygnały z poszczególnych mikrofonów nie obciążają się wzajemnie, stąd rozpatrywany obiekt MIMO o 4 wyjściach wygodniej jest potraktować jako 4 niezależne obiekty MISO, z których każdy opisuje jednoczesną zależność sygnału z odpowiedniego mikrofonu (błędu lub odniesienia) od trzech sygnałów pobudzających głośniki. Wszystkie 4 modele MISO powinny być zidentyfikowane na podstawie danych uzyskanych z jednego doświadczenia.

Aby uniknąć problemów, jakie pojawiają się przy identyfikacji modelu dla danych pochodzących z obserwacji spontanicznego przebiegu procesu lub dla eksperymentu wykonywanego w układzie zamkniętym, przyjmuje się tu, że w celu identyfikacji wspomnianych modeli przed przystąpieniem do fazy stosowania należy przeprowadzić specjalnie zaplanowany eksperyment identyfikacyjny. Eksperyment ten polegał będzie na pobudzaniu wszystkich trzech głośników (wejścia obiektu) i rejestracji sygnałów z czterech mikrofonów (wyjścia obiektu). Wykonanie eksperymentu z wykorzystaniem tego samego sprzętu i dla takiej samej konfiguracji, jaka będzie stosowana w czasie pracy układu ATH, przy niewielkiej odległości w czasie między fazami identyfikacji i aplikacji, stanowi podstawowe wymaganie dla spełnienia warunku *reprezentatywności*, por. 2.1.1. Jednocześnie, aby zapewnić spełnienie warunku *rozdzielności eksperymentatora i obiektu*, co było postulowane w 2.2.4, eksperyment przeprowadzony będzie w *układzie otwartym* (przy wyłączonym kompensatorze $W(z^{-1})$), dla specjalnie zaplanowanych sygnałów pobudzających.

Warto tu zauważyć, że w przypadku konieczności aktualizacji modelu obiektu w trakcie normalnej pracy układu ATH pojawia się istotny problem braku możliwości spełnienia warunku rozdzielenia eksperymentatora od obiektu. Przyczyną tego stanu rzeczy nie jest sprzężenie zwrotne poprzez algorytm adaptacji i filtr kompensujący – to bowiem, zgodnie z uwagami z punktu 2.2.4, jest dopuszczalne. Problemem natomiast jest jednoczesna zależność zarówno pobudzenia (poprzez tor $r(i) - W(z^{-1})$), jak i wyjścia obiektu (przez ścieżkę pierwotną) od sygnału hałasu, por. rys. 2.1. Tak więc w tej sytuacji mamy zależność wejścia i wyjścia obiektu od wspólnej dodatkowej zmiennej, co w oczywisty sposób nie spełnia warunku rozdzielności. Rozwiązaniem tego problemu może być „odsprzęgnięcie" obu sygnałów poprzez zastosowanie metod identyfikacji wykorzystujących widma wyższych rzędów, por. [38].

Rozdział 3

Planowanie eksperymentu i przygotowanie danych

Nie ulega wątpliwości, że końcowy wynik identyfikacji zależy w istotny sposób od wyników doświadczenia identyfikacyjnego. Oprócz takich czynników, jak liczba zebranych danych, ich dokładność czy poziom zakłóceń, niezwykle istotne jest także to, *w jaki sposób* eksperyment został przeprowadzony. Jednak planowanie eksperymentu stanowi ten element procedury identyfikacji, który zależy silnie zarówno od indywidualnych cech badanego procesu, jak i celu, dla którego wykonuje się identyfikację. Dlatego też wspomaganie użytkownika w tym względzie sprowadza się głównie do dostarczenia mu zestawu porad, w jaki sposób należy eksperyment przygotować, na co musi zwrócić uwagę lub czego unikać. W punkcie 3.1 omówiono ogólne zasady dotyczące właściwego przeprowadzenia eksperymentu, warunkujące poprawne wnioskowanie na podstawie uzyskanych danych. Natomiast w punkcie 3.2 przedstawiono możliwości komputerowego wspomagania w zakresie planowania eksperymentu z uwzględnieniem przede wszystkim aspektów technicznych. Szczególną uwagę zwrócono na zadanie wygenerowania odpowiedniego sygnału pobudzającego. Część z przedstawionych warunków stanowią *założenia*, których nie da się zweryfikować na drodze przeprowadzenia jakichś testów. Jednak niektóre z tych założeń mogą być testowane i dlatego w punkcie tym przedstawiono także przykłady testów zalecanych do wykonania po zebraniu danych.

Przedstawione tu zalecenia i uwagi dotyczące planowania eksperymentów identyfikacyjnych zostały zaimplementowane w dołączonej do systemu *MULTI-EDIP* aplikacji *PAIO*, której celem jest wspomaganie użytkownika w zakresie wykonywania eksperymentów przy współpracy ze specjalizowanym układem precyzyjnych wejść-wyjść analogowych *PAIO* (ang. *Precision Analog Input-Output*), por. [101].

Ponieważ zwykle dane zebrane w trakcie dowolnego eksperymentu identyfikacyjnego nie powinny być bezpośrednio stosowane do celów identyfikacji, lecz powinny zostać odpowiednio *przygotowane*, stąd w punkcie 3.3 omówiono także zadanie przygotowania danych.

Na zakończenie rozdziału przedstawiono przykład planowania eksperymentu i przygotowania danych dla identyfikacji omówionego w 2.3 obiektu elektroakustycznego.

3.1. Podstawowe wymagania

Wykonanie odpowiedniego doświadczenia może stwarzać różne problemy natury technologicznej czy ekonomicznej, a w przypadku procesów przemysłowych może wymagać uzgodnień z obsługą procesu. Stąd z powodu występujących ograniczeń nie zawsze da się zaprojektować eksperyment wyłącznie pod kątem celów identyfikacji. Jeśli np. ze względu na koszty doświadczenie musi być przeprowadzone w trakcie normalnej eksploatacji procesu, to wtedy dostępne są tylko pomiary sygnałów wejściowych i wyjściowych, bez możliwości jego specjalnego pobudzania. W przypadkach skrajnych dostępne mogą być jedynie zmienne zależne od tych, które powinny być użyte do modelowania, co bywa regułą w identyfikacji procesów ekonomicznych, socjologicznych, biologicznych itp. Jako przykład może tu posłużyć identyfikacja odpowiedzi w procesie oddychania u pacjentów z przekrwieniem serca [138], gdzie pobudzenie generowane było w postaci zmian zawartości CO_2 w powietrzu wdychanym, natomiast odpowiedź opisującą przewietrzanie płuc obliczano na podstawie mierzonego ciśnienia oraz składu wydychanego powietrza, zgodnie z zależnością podaną przez fizjologów. Czasem przeprowadzenie odpowiedniego doświadczenia może być nie tylko kosztowne czy trudne, ale wręcz nieetyczne. Jedyne, co pozostaje, to pomiar zmian wielkości tak jak się pojawiają, albo korzystanie z danych historycznych.

W takiej sytuacji powstaje oczywisty problem, w jakim stopniu uzyskane modele są *reprezentatywne* i odzwierciedlają rzeczywiste własności procesu. Teoretycznie rzecz biorąc, zbieranie danych bez znajomości źródła generującego pobudzenie trudno jest uznać za poprawnie przeprowadzony eksperyment, ponieważ wymaga to przyjęcia pewnych *założeń* odnośnie do tego źródła, które właściwie są nieweryfikowalne. Jednak w praktyce argument, że eksperyment nie dostarczył danych pozwalających na poprawną identyfikację procesu może nie być uważany za odpowiednio silny, aby przekreślić sens dalszego ich przetwarzania. Wtedy jednak może się pojawić pokusa, aby kontynuować identyfikację pomimo braku pewności co do prawdziwości przyjętych założeń. Jeżeli tego typu wstępne założenia nie będą świadomie określone, to może to w oczywisty sposób prowadzić do powstania różnego rodzaju błędów, które omówiono bliżej w punkcie 5.2.2.

Ponieważ eksperyment identyfikacyjny został opisany przez trójkę (**R, S, T**), por. 2.2.3, więc jego planowanie obejmuje określenie warunków jego prowadzenia, tj. **R** i **T**. Aby uzyskany na podstawie wnioskowania statystycznego model był *reprezentatywny*, tj. by w czasie stosowania T dla wszystkich możliwych R obiekt **S** mógł być opisany za pomocą *modelu*, który otrzymano na podstawie eksperymentu przeprowadzonego w czasie **T** przez

„eksperymentatora" **R**, spełnione muszą być dwa podstawowe wymagania, por. [17]:

- obiekt musi być odpowiednio pobudzony,
- eksperyment musi być powtarzalny.

Poniżej przedstawiono dyskuję tych warunków.

Warunek dostatecznego pobudzenia

Warunek ten stawia wymagania na zakres wartości, jakie pobudzenie u może przyjmować w trakcie eksperymentu, jak i na jego zmienność. Niech D oznacza przestrzeń wszystkich możliwych wartości (u, y), jakie mogą wystąpić na etapie stosowania, tj. dla których $p(u, y \mid R, \mathbf{S}, T) > 0$. Podstawowe wymaganie dotyczące pobudzenia jest takie, aby dla wszystkich $(u, y) \in D$ także w czasie eksperymentu identyfikacyjnego było $p(u, y \mid \mathbf{R}, \mathbf{S}, \mathbf{T}) > 0$, czyli musi istnieć dodatnie prawdopodobieństwo pojawienia się takiego pobudzenia u, które mogłoby zostać wygenerowane przez R na etapie stosowania.

Jednak w przypadku identyfikacji modeli dynamicznych interesuje nas $p(U_N, Y_N \mid R, \mathbf{S}, T)$, por. 2.2.3. Oznacza to, że w celu dostarczenia informacji o dynamice obiektu sygnał wejściowy u musi także odpowiednio zmieniać się w czasie, co można zapisać jako wymaganie, by w trakcie eksperymentu $p[u(i) \mid U_{i-1}, Y_i, \mathbf{R}, \mathbf{S}, \mathbf{T}] > 0$ dla wszystkich sekwencji U_{i-1} oraz Y_i, jakie mogłyby się pojawić na etapie stosowania. Warunek ten bywa nazywany też *warunkiem nieustającego pobudzenia* (ang. *Persistent Excitation*, *PE*), por. np. [79, 107]. Zamiast definiować wymaganie na zmienność pobudzenia w czasie za pomocą gęstości prawdopodobieństwa, wygodniej jest zdefiniować warunek *PE* w dziedzinie częstotliwości. Oznacza to, że w trakcie eksperymentu obiekt musi być pobudzany w zakresie takich częstotliwości, dla jakich będzie pobudzany na etapie stosowania.

Warunek *PE* jest związany z wymaganiem zapewnienia odpowiedniego uwarunkowania w zadaniu estymacji parametrów dla wybranej struktury modelu, por. 5.4.2. Jest to szczególnie istotne w przypadku, gdy u jest wektorem, gdyż wymagany jest wtedy także brak korelacji między elementami u.

Jeżeli obszar dopuszczalnych sygnałów D^U jest ograniczony, w szczególności możliwe jest wystąpienie tylko jednej sekwencji U_N, to wtedy nie ma możliwości wyboru sygnału pobudzającego, a więc mamy do czynienia z eksperymentem biernym. W takim przypadku jedyne, co pozostaje, to sprawdzenie, czy to pobudzenie spełnia odpowiednie wymagania.

Warunek powtarzalności

Na ogół etapy identyfikacji i stosowania są rozdzielone, a więc przedziały czasu **T** i T są z natury różne. Ponieważ zwykle nic pewnego nie można powiedzieć na temat tego, jak obiekt

będzie się zachowywał po okresie obserwacji, np. czy jego własności nie ulegną zmianie, stąd staje się oczywiste, że potrzebna jest dodatkowa wiedza, na podstawie której można będzie założyć, że obiekt pozostanie taki sam, przynajmniej w przypadku powtórzenia eksperymentu. Własność tę można nazwać *powtarzalnością*. Stanowi ona jedną z fundamentalnych zasad nauk przyrodniczych, ponieważ wszelkie niepowtarzalne jednostkowe wydarzenia nie mają dla nauki żadnego znaczenia, por. [116].

Warunek *powtarzalności* oznacza, że dla wszystkich $(u, y) \in D$ rozkład danych $(\mathbf{u}, \mathbf{y})$ powinien być taki sam, tzn. $p(u, y \mid \mathbf{R}, \mathbf{S}, \mathbf{T}) = p(u, y \mid \mathbf{R}, \mathbf{S}, T)$ dla wszystkich powtórzeń eksperymentu w dowolnym czasie T. Najprostszym sposobem testowania *powtarzalności* jest powtarzanie eksperymentu. Jednak w przypadku gdy identyfikacja może być wykonana na podstawie tylko jednego doświadczenia (co często bywa regułą), własność *powtarzalności* stanowi *założenie*, którego prawdziwość powinna być potwierdzona przez osobę prowadzącą doświadczenie.

3.2. Komputerowe wspomaganie wykonania eksperymentu

Jak już wspomniano, wymagania dla poprawnego wykonania eksperymentu zależą silnie zarówno od indywidualnych cech analizowanego procesu, jak i celu, dla którego wykonuje się identyfikację, stąd bardzo trudno jest przedstawić w sposób formalny pewne ogólne reguły, które zaprogramowane w komputerze pozwoliłyby na wykonanie eksperymentu w sposób możliwie zautomatyzowany. Dotyczy to szczególnie przedstawionych wyżej ogólnych warunków poprawnego przeprowadzenia eksperymentu, jako że w istocie stanowią one pewne *założenia*, które trudno jest zweryfikować za pomocą jakiegoś oprogramowania.

Inne podstawowe wymaganie wobec eksperymentu, takie jak *warunek rozdzielności* eksperymentatora i obiektu, por. 2.2.4, może być automatycznie zapewnione tylko w przypadku realizacji eksperymentu w *układzie otwartym*. W sytuacji wykonywania eksperymentu w *układzie zamkniętym* albo dla *pobudzenia spontanicznego* system wspomagania nie tylko nie jest w stanie zapewnić spełnienia wspomnianego warunku rozdzielenia, ale nawet nie jest w stanie tego sprawdzić na podstawie uzyskanych danych. Albowiem jedyne, co można testować, to korelacja między sygnałami, a ona może być spowodowana przez istnienie ukrytej zależności od dodatkowej wspólnej zmiennej, por. 5.2.2. W takim przypadku na użytkowniku spoczywa cała odpowiedzialność za sprawdzenie, że warunek ten jest rzeczywiście spełniony. Podstawową formą komputerowego wspomagania osoby prowadzącej identyfikację może być więc tu jedynie podanie pewnych ogólnych porad wynikających z wymagań warunkujących poprawne wnioskowanie na podstawie zebranych danych.

Jednak oprócz takich ogólnych wymagań wobec eksperymentu, z planowaniem eksperymentu wiążą się także decyzje dotyczące zagadnień, które są natury bardziej technicznej. Poza określeniem wejść u i wyjść y procesu należą tu:

- sposób ich pomiaru, w tym określenie punktu pracy, dla którego będzie identyfikowany model procesu oraz dopuszczalnych zakresów zmian sygnałów;
- określenie połączenia między procesem a komputerem (eksperymentatorem), w tym sposobu akwizycji danych;
- określenie sposobu pobudzania obiektu, w tym zaprojektowanie (o ile to możliwe) sygnału pobudzającego;
- dobór okresu próbkowania.

Ponieważ pierwszy element powyższej listy jest ściśle związany z procesem, który ma być identyfikowany, oraz ze sposobem wykorzystania modelu, więc w praktyce także i w tym zakresie główną formą wspomagania osoby prowadzącej identyfikację może być jedynie podanie pewnych ogólnych zaleceń, których zastosowanie umożliwia spełnienie podstawowych warunków dla poprawnego wykonania eksperymentu. Mogą to być porady typu:

- Eksperyment identyfikacyjny powinien być wykonany w warunkach zbliżonych do tych, w jakich będzie wykorzystywany otrzymany model.

  Jest to warunek konieczny (ale niewystarczający) do spełnienia warunku *reprezentatywności*.

- Należy dążyć do zmniejszenia wpływu błędów i szumów pomiarowych na wyniki.

  Podstawowe wymagania to dopasowanie zakresu przetworników do zakresów zmienności sygnałów, dopasowanie elektrycznych własności obwodów oraz odpowiednie zabezpieczenie przed wpływem zewnętrznych pól elektromagnetycznych przy przesyłaniu sygnałów.

- Wybrać odpowiedni punkt pracy.

  Chodzi tu o taki wybór wartości sygnałów w stanie ustalonym, aby obiekt znajdował się w odpowiednim punkcie jego charakterystyki statycznej, a model dynamiczny opisywał zachowanie obiektu dla odchyłek od tego punktu. Powinien to być oczywiście taki punkt pracy, w którym obiekt będzie się znajdował na etapie stosowania modelu. W szczególnym przypadku punkt pracy może być określony przez zerową wartość sygnału wejściowego. Jednak w niektórych przypadkach utrzymanie obiektu w punkcie pracy może być trudne i może wręcz się wiązać z koniecznością prowadzenia eksperymentu w układzie *ze sprzężeniem zwrotnym*, zapewniającym spełnienie tego wymagania. W takim przypadku, aby możliwe było wyróżnienie obiektu od eksperymentatora, sprzężenie to dozwolone jest tylko od wartości wyjściowej y i powinno być realizowane przez ten sam komputer, por. 2.2.4. Ponadto sprzężenie to musi być takie, aby zadanie wyznaczenia parametrów nie stało się osobliwe, najlepiej aby było ono nieliniowe lub niestacjonarne, a przynajmniej odpowiedniego rzędu, albo do pętli sprzężenia zwrotnego należy wprowadzić dodatkowy sygnał pobudzający, por. np. [43, 62].

- Eksperyment powinien być prowadzony w taki sposób, by usunąć wpływ warunków początkowych, w szczególności dochodzenia do punktu pracy.
- Zakres zmienności sygnałów oraz przetworników należy dobrać tak, aby uniknąć wchodzenia w obszar charakterystyk nieliniowych, np. w nasycenie.

Szczególnie korzystna sytuacja występuje wtedy, gdy eksperyment identyfikacyjny może być wykonany przy użyciu odpowiedniego wyposażenia technicznego, które będzie później wykorzystane także na etapie stosowania. W kolejnych podpunktach omówiono zagadnienia sprzężenia procesu z komputerem pełniącym rolę eksperymentatora, wyboru pobudzenia oraz okresu próbkowania.

3.2.1. Sprzężenie procesu z komputerem

Jeżeli eksperyment identyfikacyjny będzie wykonany z wykorzystaniem odpowiednich układów akwizycji danych i generowania pobudzenia, to zdecydowanie zaleca się, aby był to *ten sam* sprzęt, który później będzie wykorzystany na etapie stosowania. Albowiem wszystkie elementy torów przesyłania sygnałów, takie jak *układ wejść-wyjść analogowych* czy wszelkiego rodzaju *filtry*, powinny być traktowane jako część procesu **P**, a nie eksperymentatora **R**, por. 2.2.3. Stąd spełnienie warunku *reprezentatywności* będzie możliwe tylko wtedy, gdy będą one *takie same* na etapach identyfikacji i stosowania. Jeżeli dodatkowo możliwe będzie wykonanie eksperymentu w taki sposób, aby pobudzenie było z góry ustalonym ciągiem pobudzającym, to jest to najbardziej pożądany przypadek identyfikacji w *układzie otwartym*, zapewniający spełnienie warunku *rozdzielności* eksperymentatora i obiektu.

Zwykle w sytuacji najczęściej spotykanych celów, takich jak sterowanie i predykcja, mamy do czynienia z procesami ciągłymi w czasie, ale próbkowanymi, które należy opisać za pomocą modeli dyskretnych w czasie. Zadaniem układu *wejść analogowych* jest zamiana sygnału ciągłego na sygnał dyskretny w czasie (*próbkowanie*) oraz zamiana sygnału analogowego

(napięciowego lub prądowego) pochodzącego z przetwornika pomiarowego na sygnał cyfrowy (*kwantowanie*). Obie czynności mogą stanowić źródła błędów: *próbkowania* i *kwantowania*.

Operacja próbkowania wykonywana jest przez układ próbkująco-pamiętający. W celu identyfikacji modeli dynamicznych dyskretnych w czasie dla procesów ciągłych z próbkowaniem sygnały muszą być próbkowane ze *stałą częstotliwością*, przy czym powinny być one próbkowane *synchronicznie*. Jest to ważne wymaganie, które w praktyce może być trudne do spełnienia. Błędy powodowane fluktuacją chwil próbkowania nazywane są migotaniem i wpływają na kształt widma gęstości mocy, szczególnie dla wyższych częstotliwości [110]. W przypadku niespełnienia drugiego warunku fakt przesunięcia fazowego między próbkowaniem kolejnych sygnałów powinien być uwzględniony na etapie stosowania zidentyfikowanego modelu, np. poprzez zapewnienie istnienia takiego samego przesunięcia. Stosowane są także metody przetwarzania sygnałów próbkowanych ze zmienną częstotliwością, głównie w celu uniknięcia zjawiska stroboskopowego, por. np. [12, 87].

Operację kwantowania wykonuje przetwornik analogowo-cyfrowy, o którego dokładności decyduje przede wszystkim jego *rozdzielczość*. Aby w jak najlepszy sposób wykorzystać rozdzielczość przetwornika, zakres sygnału wejściowego (napięciowego lub prądowego) przetwornika analogowo-cyfrowego (A/C) powinien być tak dobrany, aby był on jak najbliższy poziomowi sygnału wyjściowego z przetwornika pomiarowego, a jednocześnie aby zapewniał niewchodzenie w zakres nieliniowości (nasycenie). Jeśli rozdzielczość przetwornika jest nie gorsza niż 10 bitów, to błąd kwantowania można w zasadzie pominąć, por. [105, 110].

Tu jednak mogą się pojawić pewne problemy gdy poszukiwany jest model procesu dla niewielkich odchyłek od punktu pracy, a przetwornik A/C pokrywa pełny zakres zmian mierzonego sygnału. W takiej sytuacji bowiem, po odjęciu wartości odpowiadających ustalonemu punktowi pracy (tzw. *offset*) od wartości zmierzonych, otrzymuje się wartości pomiarów z rozdzielczością zaledwie kilku bitów (jedna, dwie cyfry znaczące). To oznacza, że przy modelowaniu należy uwzględnić błędy zaokrągleń, które mogą być traktowane jako biały szum dodawany do pomiarów, co w przypadku identyfikacji modeli parametrycznych można rozwiązać stosując specjalne metody, por. [59].

Układ *wyjść analogowych* służy do zamiany sygnału cyfrowego dyskretnego w czasie na sygnał (napięciowy lub prądowy) ciągły w czasie, wprowadzany na wejście elementu wykonawczego. Pełni więc rolę ekstrapolatora zerowego rzędu. Tego typu generowanie pobudzenia jest wymagane, jeżeli na etapie stosowania modelu obiekt będzie sterowany komputerowo także z użyciem ekstrapolatora zerowego rzędu.

Sytuacja się komplikuje, jeżeli pobudzenie nie jest wprowadzane na wejście obiektu poprzez układ wyjść analogowych, ale generowane w taki sposób, że możliwy jest tylko pomiar

sygnału wejściowego. W tym przypadku bowiem pomiar ten jest także obarczony jakimś błędem, który należałoby modelować w postaci błędu dodawanego do wejścia. Podejście takie, aczkolwiek proponowane np. w [115], jest jednak niewygodne i dlatego wyjściem z tego dylematu może być założenie, że to, co jest mierzone jako sygnał wejściowy, stanowi *dokładną wartość na wejściu*, natomiast błąd tego pomiaru jest w domyśle przenoszony na wyjście i traktowany jako *zakłócenie sprowadzone na wyjście*, podobnie do innych źródeł zakłóceń, por. 4.2.

Jednak to nie rozwiązuje wszystkich problemów, ponieważ w celu poprawnej identyfikacji wymagane jest, aby widmo sygnału pobudzającego było ograniczone. Jeżeli więc w sygnale tym nie występują składowe powyżej pewnej częstotliwości granicznej ω_g, to przyjmując okres próbkowania $\tau_p \leq \pi/\omega_g$ i wprowadzając odpowiednie filtry antystroboskopowe na wejściu i wyjściu można uzyskać dane bez utraty informacji istotnej z punktu widzenia modelowania procesu [79].

Inny problem występuje wtedy, gdy danymi wziętymi do identyfikacji nie są bezpośrednio wartości pomiarów, ale zmienne uzyskane na drodze jakiegoś przetwarzania tych pomiarów. W takim przypadku niezwykle istotne jest to, aby sposób tego przetwarzania był dokładnie taki sam dla wszystkich danych wziętych do identyfikacji, a także na etapie stosowania modelu.

Przedstawione tu uwagi mogą być w systemie wspomagania identyfikacji zapisane w postaci zestawienia porad dla użytkownika, w jaki sposób należy wykonać sprzężenie komputera i procesu. Pewną pomocą dla niego może być dostarczenie gotowego sprzętu oraz aplikacji do obsługi tego sprzętu, czego przykładem jest system *MULTI-EDIP*, współpracujący z układem wejść-wyjść analogowych *PAIO*, por. [101].

3.2.2. Sygnał pobudzający

W punkcie 3.1 omówiono podstawowe wymagania wobec sygnału pobudzającego proces (o ile występuje). Jednym z najpowszechniejszych problemów pojawiających się, gdy niedoświadczony użytkownik przystępuje do identyfikacji, jest brak odpowiedniego pobudzenia wymaganego do identyfikacji wybranej struktury modelu.

Jeśli istnieje możliwość przeprowadzenia doświadczeń *czynnych*, tj. takich, w których istnieje możliwość generacji specjalnie zaprojektowanych sygnałów wejściowych, to jest to sytuacja idealna z punktu widzenia identyfikacji. Niestety, nie zawsze jest to możliwe i wtedy jedyne, co pozostaje, to testowanie poprawności zebranych danych, np. czy sygnał wejściowy stanowi pobudzenie odpowiedniego stopnia, por. 3.2.4. Jeżeli okaże się, że dane nie spełniają odpowiednich testów, to właściwie pozostaje powtórzenie doświadczenia. Czasem może tu pomóc stosowanie specjalnych technik przetwarzania, takich jak np. faktoryzacja *QR* [35].

W sytuacji gdy można zaprojektować sygnał pobudzający przeznaczony specjalnie na potrzeby identyfikacji, do ustalenia są dwa parametry projektowe dla tego sygnału:

- zakres wartości, jakie może sygnał przyjmować (amplituda),
- przebieg sygnału w czasie.

Zalecenia dotyczące wyboru amplitudy wynikają z właściwości procesu. W [62] przedstawiono je w ogólnych zarysach.

Jeżeli natomiast chodzi o zmienność sygnału w czasie, to bardzo pomocne może być tu dostarczenie użytkownikowi możliwie bogatego zestawu generatorów różnych sygnałów pobudzających, tak aby nie musiał samodzielnie programować odpowiednich funkcji. Powinny to być zarówno proste sygnały deterministyczne, umożliwiające przeprowadzenie wstępnych eksperymentów identyfikacyjnych dla pozyskania informacji *a priori* o własnościach procesu (np. w celu określenia okresu próbkowania), jak i sygnały złożone, spełniające warunki *PE* dla pobudzenia obiektu, por. 3.1. Sposoby generacji sygnałów przeznaczonych dla różnych celów identyfikacji omówiono w [39], natomiast w [101] przedstawiono przykładowy zestaw takich sygnałów, pokrywający zdecydowaną większość zastosowań i ułatwiający użytkownikowi systemu komputerowego wspomagania identyfikacji wygenerowanie sygnału pobudzającego odpowiedniego dla jego celu.

Odpowiedni przebieg sygnału w czasie można także uzyskać na drodze filtracji sygnałów pseudolosowych przez specjalnie zaprojektowany filtr cyfrowy, co umożliwia np. uzyskanie wymaganych własności częstotliwościowych sygnału pobudzającego.

Ze względu na wiele zalet warto jest w systemie wspomagania identyfikacji ukierunkowywać użytkownika na stosowanie sygnałów wielosinusoidalnych okresowych. Konstruując taki sygnał, częstotliwości poszczególnych sinusoid dobiera się tak, aby zadany okres sygnału wielosinusoidalnego był całkowitą wielokrotnością wszystkich jego składowych. Jeżeli liczba tych sinusoid będzie równa połowie liczby próbek w okresie (maksymalna wartość), to sygnał taki jest *PE* stopnia równego liczbie tych próbek. Stosowanie do pobudzenia obiektu tego typu sygnałów (pseudo) losowych umożliwia uśrednianie uzyskanych danych dla kolejnych okresów. Pozwala to na zmniejszenie liczby danych stosowanych do aproksymacji modelu przy jednoczesnym zmniejszeniu stosunku szumu do sygnału, lub na ominięciu problemów związanych z wykonywaniem eksperymentu w układzie zamkniętym, por. [30]. W przypadku wykorzystywania transformacji Fouriera w przetwarzaniu danych sygnały te zapobiegają powstawaniu zjawiska wycieku widma, por. np. [11]. Ponadto sygnały wielosinusoidalne losowe ułatwiają identyfikację modelu liniowego w sytuacji występowania nieliniowych zakłóceń, por. [122]. Aby jednocześnie uzyskać zwiększenie stosunku sygnału do

szumu przy ograniczonym zakresie zmian wartości, można tu wykorzystać sygnały wielosinusoidalne, w których fazy generowane są metodą Schroedera zmniejszającą współczynnik szczytu (definiowany jako stosunek różnicy wartości ekstremalnych sygnału do jego odchylenia standardowego) na podstawie specjalnej heurystyki, por. [123], lub metodą Ouderaa, por. [111].

Problem odpowiedniego pobudzenia komplikuje się w sytuacji, gdy w obiekcie występuje więcej wejść. W takim przypadku bowiem sygnały pobudzające na poszczególnych wejściach powinny być takie, aby możliwa była identyfikacja poszczególnych torów sterowania modelu *MISO* lub *MIMO* w sposób wzajemnie niezależny. Aby to osiągnąć, można zastosować dwa rozwiązania:

- powtarzać eksperyment, wprowadzając odpowiednie pobudzenie na kolejne wejście i utrzymując jednocześnie na pozostałych wejściach stały sygnał, który odpowiada punktowi pracy. Jest to pracochłonne i kosztowne, ale umożliwia identyfikację poszczególnych torów sterowania niezależnie od pozostałych;
- zastosować na poszczególnych wejściach sygnały nieskorelowane, najlepiej ortogonalne, dzięki czemu dane uzyskane w jednym doświadczeniu pozwolą na identyfikację wszystkich torów sterowania niezależnie między sobą.

To drugie rozwiązanie jest szczególnie zalecane, gdyż oszczędza czas i zmniejsza koszty. Jako sygnały pobudzające spełniające powyższe wymagania najwygodniej jest zastosować *wektorowe sygnały wielosinusoidalne ortogonalne* (*WSWO*). Ich ortogonalność przejawia się w tym, że na poszczególnych wejściach występują sygnały posiadające składowe sinusoidalne o częstotliwościach nie występujących na innych wejściach. Sposób generacji takich sygnałów przedstawiono w [29], a możliwości ich wykorzystania do identyfikacji obiektów *MISO* i korzyści płynące z ich stosowania omówiono w [28] i [32]. Przykład wykorzystania własności sygnałów WSWO do identyfikacji obiektu elektroakustycznego pokazano w rozdziale 7.

Inną możliwość uzyskania sygnałów pobudzających nieskorelowanych stanowią sekwencje wektorowych ciągów pseudolosowych binarnych (*PRBS*) generowanych z wykorzystaniem algorytmu powtarzania z inwersją (ang. *inverse–repeated*), por. [39].

W literaturze proponowane są także metody optymalizacji sygnału pobudzającego ze względu na maksymalizację pewnej miary dokładności estymacji, por. np. [34, 41]. Jednak ich wadą jest konieczność posiadania odpowiedniej wiedzy *a priori* na temat parametrów procesu, co zwykle możliwe jest tylko wtedy, gdy parametry te mają fizyczną interpretację lub istnieją ich wcześniejsze oszacowania w literaturze. Inne możliwe rozwiązanie polega na iteracyjnym poprawianiu eksperymentu na podstawie wyników identyfikacji modelu uzyskanych

w poprzedniej iteracji. Ze względów ekonomicznych i praktycznych tego typu podejście może być trudne do realizacji, stąd nie ma potrzeby implementacji tych metod w systemie wspomagania identyfikacji.

3.2.3. Próbkowanie

Wybór częstotliwości próbkowania może być w naturalny sposób narzucony przez charakter identyfikowanego procesu oraz zastosowanie modelu. Dotyczy to szczególnie takich dziedzin nauki, jak ekonomia, socjologia, biologia itp. oraz celów związanych głównie z predykcją lub opisem własności procesu. W przypadku identyfikacji procesów dla celów sterowania (w szczególności procesów przemysłowych lub związanych z zastosowaniami technicznymi), gdzie poszukiwany jest model dyskretny w czasie dla procesów ciągłych z próbkowaniem, wybór częstotliwości próbkowania stanowi jeden z najistotniejszych parametrów projektowych dla eksperymentu identyfikacyjnego.

W zasadzie częstotliwość próbkowania na etapie identyfikacji powinna być taka sama jak na etapie stosowania, chociaż czasami jest możliwe przeliczenie parametrów dla zmiany częstotliwości, np. przy stałym sygnale wejściowym w przedziale próbkowania. Oprócz tego wybór tej częstotliwości podlega także pewnym ograniczeniom. Znane jest powszechnie ograniczenie dolne wynikające z twierdzenia Shannona-Kotielnikowa: $f_p = 1/\tau_p \geq 2f_{max}$, gdzie f_p i τ_p oznaczają odpowiednio częstotliwość i okres próbkowania, a f_{max} maksymalną częstotliwość występującą w widmie sygnału próbkowanego, por. np. [126].

Ponieważ jednak często widmo sygnału próbkowanego zawiera szereg wyższych częstotliwości, nieistotnych z punktu widzenia zastosowania modelu, a wynikających z obecności w sygnałach szumów pomiarowych i zakłóceń, składowe te powinny być *odfiltrowane przed próbkowaniem*. Zapobiega to powstawaniu zjawiska polegającego na nakładaniu się składowych widma powyżej częstotliwości Nyquista na składowe niskoczęstotliwościowe, zwanego zjawiskiem *stroboskopowym* albo *przeinaczania widma* (*aliasing*), por. np. [12, 87]. Tak więc pomiędzy źródłem sygnału (przetwornikiem pomiarowym) a przetwornikiem analogowo-cyfrowym w układzie wejść analogowych (ściśle biorąc, układem próbkująco-pamiętającym) powinien być wprowadzony *analogowy filtr dolnoprzepustowy* (antystroboskopowy) o charakterystyce częstotliwościowej zapewniającej odpowiednie tłumienie składowych widma powyżej częstotliwość *Nyquista*.

Większość dostępnych kart analogowych do komputerów nie zawiera takich filtrów, pozostawiając ich konstrukcję użytkownikom. Jednak zaprojektowanie odpowiedniego filtru może być szczególnie trudne w sytuacji, gdy przewiduje się pracę układu dla różnych częstotliwości próbkowania, a więc gdy konieczne jest stosowanie filtru analogowego

o nastawialnej częstotliwości odcięcia. Specjalizowany układ wejść-wyjść analogowych, z którym współpracuje system *MULTI-EDIP*, został wyposażony w odpowiedni zestaw filtrów antystroboskopowych, przestrajalnych w szerokim zakresie częstotliwości, por. [101]. Jest to jednak rozwiązanie unikalne.

Jeżeli sygnał wejściowy u jest generowany poprzez ekstrapolator zerowego rzędu, to filtr taki powinien być stosowany *tylko* dla sygnału wyjściowego y i wtedy powinien być traktowany jako element procesu. W celu spełnienia warunku *reprezentatywności* musi być to *taki sam* filtr, jaki będzie użyty na etapie stosowania. Tak więc przewidując użycie filtru antystroboskopowego w trakcie wykonywania eksperymentu identyfikacyjnego trzeba od razu mieć na względzie aplikację systemu, w szczególności wpływ dynamiki tego filtru na pogorszenie dynamiki odpowiedzi całego obiektu (objawiający się zwiększeniem rzędu i pojawieniem się dodatkowych opóźnień). Rozwiązaniem tego problemu nie jest jednoczesna filtracja sygnałów wejściowego i wyjściowego za pomocą filtrów o takiej samej charakterystyce i zakładanie, że wtedy nadal będzie możliwa prawidłowa identyfikacja dynamicznych własności wejściowo-wyjściowych. Wynika to stąd, że są to filtry *analogowe*, które filtrują sygnały *przed* próbkowaniem. Jeżeli zebrane ciągi próbek zostaną poddane filtracji *cyfrowej*, przy czym filtry te będą miały *dokładnie taką samą* transmitancję, to na podstawie ciągów danych po filtracji można nadal uzyskać *poprawny* model dynamiczny dla zależności między wejściem a wyjściem obiektu. Natomiast w tym przypadku filtracja analogowa sygnału wprowadzanego na wejście obiektu z ekstrapolatora zerowego rzędu spowoduje, że rzeczywisty sygnał wejściowy poddany próbkowaniu nie będzie spełniał podstawowego wymagania, jakim jest stałość sygnału pomiędzy chwilami próbkowania, a więc niemożliwe będzie uzyskanie poprawnego modelu obiektu, por. [105].

Oprócz wspomnianego ograniczenia dolnego na częstotliwość próbkowania istnieje także ograniczenie górne, wynikające z jednej strony z konieczności zapewnienia odpowiedniego uwarunkowania numerycznego zadania aproksymacji, a z drugiej ze względów praktycznych, por. [62]. Albowiem aby zarejestrować przebiegi sygnałów w odpowiednio długim przedziale czasu, wystarczającym do zaobserwowania istotnych własności dynamicznych badanego procesu, trzeba by zgromadzić, a następnie przetwarzać, olbrzymią liczbę pomiarów. W przeciwnym przypadku, przy ograniczonej zdolności magazynowania danych, trzeba by przerwać eksperyment, zanim w obiekcie nastąpiłyby odpowiednie reakcje na zastosowany sygnał pobudzający. Ponadto zbytnie skracanie okresu próbkowania może prowadzić do problemów z estymacją niektórych parametrów, por. [104].

Stąd w podręcznikach podaje się pewne heurystyczne reguły wyboru odpowiedniego okresu próbkowania, por. np. [48, 62, 79], zapewniające na ogół uzyskanie dobrych wyników.

Reguły takie powinny być także włączone do komputerowego systemu wspomagania, o ile w systemie tym przewidziano możliwość prowadzenia eksperymentu identyfikacyjnego. Jednak ich stosowanie wymaga pewnej wiedzy *a priori* o identyfikowanym procesie (np. informacji o dominujących stałych czasowych), stąd albo wiedzę taką musi dostarczyć użytkownik systemu, albo system musi umożliwić przeprowadzenie wstępnego eksperymentu, np. w postaci badania odpowiedzi skokowej obiektu.

Trzeba także użytkownikowi systemu uświadomić, że jeżeli dane pomiarowe zostały zebrane ze zbyt dużą częstotliwością, to w przypadku stosowania ekstrapolatora nie można w sposób sztuczny zwiększać okresu próbkowania, wybierając do przetwarzania tylko co którąś próbkę, o ile pomiędzy tymi próbkami nie zapewniono stałości sygnału pobudzającego obiekt. Stąd jeżeli z góry przewiduje się, że na etapie *przygotowania danych*, por. 3.3, będzie możliwość ich edycji w celu zwiększenia okresu próbkowania poprzez wybór do przetwarzania tylko co którejś próbki, to wtedy eksperyment identyfikacyjny powinien być prowadzony w taki sposób, by sygnał wejściowy zmieniał się tylko co określoną liczbę próbek. W takim przypadku oprócz okresu próbkowania τ_p można wyróżnić także okres sterowania τ_s, tzn. okres będący wielokrotnością okresu próbkowania, w czasie którego sygnał pobudzający ma stałą wartość. Rozwiązanie takie zaproponowano w systemie *MULTI-EDIP*, por. [101]. Jednak wtedy trzeba pamiętać o tym, że w ten sposób ogranicza się zakres częstotliwości, dla których obiekt jest pobudzany, por. [62].

3.2.4. Testowanie poprawności uzyskanych danych

Część z przedstawionych warunków dla poprawnego wykonania eksperymentu identyfikacyjnego stanowi *założenia*, których nie da się zweryfikować na drodze testowania zebranych danych, jak np. założenie o *rozdzielności* eksperymentatora i obiektu. Podobnie na podstawie samych danych nie jest możliwe wykrycie istnienia *zmiennej ukrytej*. Natomiast założenie o *powtarzalności* może być w ograniczonym zakresie testowane na drodze powtarzania eksperymentu. W szczególny sposób można to zrobić stosując do testowania modelu zbiór danych pochodzących z innego eksperymentu niż dane wzięte do estymacji, por. 5.4.1.

Jednak część wymagań wobec eksperymentu może być sprawdzona na podstawie samych danych. Stąd zgodnie z zasadą, że to wszystko, co może być testowane, powinno być testowane jeszcze przed rozpoczęciem procedury identyfikacji, niezbędne jest uzupełnienie komputerowego systemu wspomagania identyfikacji o tego typu testy, aby ustrzec się niepotrzebnych błędów.

Podstawowym takim testem jest sprawdzenie, czy w ciągu danych nie występują tzw. *wartości skrajne*, zwane też *błędami grubymi* (ang. *outliers*), tj. próbki, których bezwzględne odchylenia od wartości średniej są kilkakrotnie większe od odchylenia standardowego ciągu danych. Nawet kilka wartości skrajnych może wpływać na oceny parametrów modelu w sposób bardzo znaczący, por. [79]. Dlatego do systemu wspomagania identyfikacji powinna być włączona procedura sprawdzania obecności takich błędów w próbkach i ich ewentualnej korekcji, np. na drodze *ekstrapolacji* lub *interpolacji liniowej*, por. [101]. Z oczywistych względów procedury takiej nie stosuje się dla sygnałów pobudzających, lecz jedynie dla wyjść procesu.

Możliwe jest także, aczkolwiek w ograniczonym sensie, sprawdzanie poprawności okresu próbkowania. Ponieważ wybór zbyt małej częstotliwości próbkowania może prowadzić do powstania efektu stroboskopowego, stąd badając periodogram wygładzony w dziedzinie częstotliwości można wyznaczyć wskaźnik *błędu okresu próbkowania*, zdefiniowany jako procentowy stosunek wartości wygładzonego periodogramu dla częstotliwości Nyquista $\omega\tau_p = \pi$ do maksymalnej wartości tego periodogramu, por. [101]. Formalnie rzecz biorąc, na podstawie próbek sygnału nie można orzec, czy zjawisko stroboskopowe miało miejsce w czasie próbkowania czy nie. Jednak można się tu posłużyć pewnym rozumowaniem, które opiera się na założeniu, że widmo sygnału jest *ciągłe* i *ograniczone*. Jeżeli założenie to jest spełnione, to w sytuacji, gdy widmo sygnału po próbkowaniu ma dla częstotliwości Nyquista wartość bliską zeru, można *domniemywać*, że zjawisko to nie zaszło. Jeżeli jednak wartość ta będzie znacząco większa od zera, to dowodzi to powstania efektu stroboskopowego, a zatem wskazuje na konieczność *powtórzenia* eksperymentu identyfikacyjnego dla odpowiednio *zwiększonej* częstotliwości próbkowania, albo zastosowania odpowiednich *filtrów antystroboskopowych*.

Oprócz przedstawionych tu testów możliwych do automatycznego wykonania bezpośrednio po wygenerowaniu danych, użytkownik systemu może uzyskać jeszcze dodatkowe wspomaganie już po wyznaczeniu modelu. Otóż w przypadku identyfikacji modelu *parametrycznego* na etapie *testowania modelu* powinno się wykonać testy *uwarunkowania równań estymatora*, por. 5.4.2. W przypadku wykrycia słabego uwarunkowania można wykonać testy pomocnicze, które wspomogą użytkownika w wykryciu tych przyczyn słabego uwarunkowania, które wynikają ze złego zaprojektowania eksperymentu, por. dodatek B. Niektóre z proponowanych tu testów pomocniczych zostały omówione także w [101].

Jednak przed przystąpieniem do identyfikacji, szczególnie modeli parametrycznych, warto jest w systemie komputerowego wspomagania identyfikacji wykonać kilka niezwykle prostych, lecz skutecznych testów badających uwarunkowanie numeryczne danych. Ponieważ osoba prowadząca identyfikację może nie zdawać sobie sprawy z istoty tego problemu, więc testy

takie powinny być wykonywane automatycznie, bez żądania ze strony użytkownika, a ich celem jest ostrzeżenie użytkownika, jeszcze przed wykonaniem estymacji parametrów modeli, że dane pomiarowe wzięte do identyfikacji nie zapewnią odpowiedniego uwarunkowania układu równań estymatora. Jest wprost zadziwiające, jak czasami najprostsze działania mogą ustrzec użytkownika, szczególnie niedoświadczonego, przed niepotrzebnymi błędami. Proponuje się tu testowanie następujących wymagań:

- Czy wszystkie ciągi danych pomiarowych są odpowiednio zmienne?

  Test ten polega na sprawdzeniu, czy odchylenie standardowe (*OS*) dowolnego ciągu nie jest mniejsze lub większe od założonej granicy, np. $OS < 10^{-6}$ lub $OS > 10^{6}$.

- Czy w którymś z ciągów danych nie występuje duża wartość średnia (tzw. *offset*)?

  Jeśli np. wartość średnia ciągu jest większa niż jego odchylenie standardowe, to można uważać, że ciąg zawiera *offset*, który powinien być usunięty.

- Czy nie występują zbyt duże różnice pomiędzy wartościami pomiarów dla poszczególnych ciągów?

  Zbyt duże różnice w wartościach danych między różnymi ciągami mogą być przyczyną słabego uwarunkowania zadania estymacji, co może prowadzić do wyników podatnych na wszelkie błędy oraz małej precyzji modelu. Można np. przyjąć, że stosunek odchyleń standardowych dla dowolnej pary przetwarzanych ciągów powinien się mieścić w granicach od *0.001* do *1000*.

W każdym przypadku pozytywnego wyniku któregoś z powyższych testów użytkownik powinien otrzymać odpowiednie ostrzeżenie o występujących defektach w danych pomiarowych. W pierwszym przypadku dany ciąg należy albo usunąć z przetwarzanych danych, albo odpowiednio go przeskalować. W drugim przypadku należy z ciągu usunąć *offset* - albo przez odjęcie wartości średniej, albo zadanej wartości stałej. W trzecim przypadku należy odpowiednio przeskalować pomiary, aby odchylenia standardowe poszczególnych ciągów były w miarę przybliżone. Działania te mogą być wykonane w ramach przygotowania danych omówionego w następnym punkcie, a o ich podjęciu decyduje użytkownik.

Jeżeli odpowiedzi na powyższe pytania są negatywne, to przed przystąpieniem do identyfikacji warto jeszcze porównać wartości funkcji autokorelacji $\rho_{vy}(l)$ obliczonej dla przesunięcia l równego 1 i 0 dla ciągów wyjściowych procesu. Jeżeli wartości te są bliskie sobie, np. $\rho_{vy}(1) > 0.9\ \rho_{vy}(0)$, to może to świadczyć o tym, że albo w ciągu tym występuje trend, który powinien zostać usunięty, por. 3.3.3, albo przyjęto zbyt krótki okres próbkowania, co

powoduje zbyt małą zmienność sygnału w czasie. W tym drugim przypadku należy albo powtórzyć eksperyment, albo wykonać operację decymacji, por. 3.3.1.

3.3. Wspomaganie w zakresie przygotowania danych

Dane surowe, czyli dane zebrane w trakcie eksperymentu identyfikacyjnego, zwykle nie nadają się do bezpośredniego wykorzystania do celów identyfikacji, ale powinny być poddane pewnemu wstępnemu przetwarzaniu, które pozwoli lepiej wykorzystać zawartą w nich informację lub usunąć pewne ich wady. Pozostawienie w danych stosowanych do identyfikacji takich wad, jak np. zakłócenia niskoczęstotliwościowe typu *płynięcie* sygnału czy zmiany *poziomu odniesienia*, albo zakłócenia o częstotliwościach większych niż te, które są istotne z punktu widzenia własności dynamicznych identyfikowanego procesu, jest często spotykanym źródłem złych wyników. Podobnie ciąg danych surowych może zawierać wiele *niepotrzebnych fragmentów*, tzn. podciągów nie zawierających informacji istotnych dla identyfikacji, które powinny być usunięte, albo może wymagać podzielenia na podciągi, które będą przetwarzane oddzielnie, w celu wyjaśnienia pewnych szczególnych zagadnień związanych ze strukturą modelu, np. zagadnień stacjonarności.

W związku z tym warto jest wyposażyć system wspomagania w odpowiednio bogaty zestaw procedur umożliwiających *edycję danych* oraz ich *wstępne przetwarzanie*. Ponieważ stosowanie algorytmów przygotowania danych zależy w istotny sposób od konkretnego zadania, więc o ich wykorzystaniu decyduje użytkownik systemu. Ważne jest natomiast to, aby przedstawić mu wagę tego problemu, a jeśli rezygnuje z ich stosowania, to by robił to świadomie. Jednocześnie w podjęciu decyzji co do stosowania pewnych technik wstępnego przetwarzania system może go wspomóc przez udostępnienie testów dodatkowych lub możliwości identyfikacji modeli deterministycznych, co omówiono w 3.3.3.

3.3.1. Edycja danych

Przez pojęcie edycji danych rozumie się tu działania mające na celu zmianę rozkładu danych w czasie, takie jak wycięcie z ciągu próbek *wybranego podciągu* lub *zmiana okresu próbkowania*. Wyboru przedziału czasu użytkownik może dokonać na podstawie analizy wzrokowej przebiegu czasowego ciągów danych lub innej wiedzy na temat przeprowadzonego eksperymentu. Ważne jest, aby w wybranym przedziale czasu występowała odpowiednia zmienność analizowanych sygnałów oraz aby przedział ten był reprezentatywny dla identyfikowanego procesu. Pewną pomocą może być przedstawienie wykresów *średniej ruchomej* (*Moving Average*, *MA*) dla zadanej szerokości okna lub *średniej przedziałowej*,

zdefiniowanej poprzez zadeklarowanie kilku przedziałów czasu, dla których wyznacza się wartości średnie, por. [101]. Wykresy te pozwalają ocenić zmienność w czasie badanego ciągu oraz wyznaczyć przedziały czasu (zakresy numerów próbek), które powinny być wzięte do przetwarzania. Wybór szerokości okna i przedziałów należy do użytkownika.

Zwiększenie okresu próbkowania o całkowitą wielokrotność wykonywane jest za pomocą operacji decymacji, a zmniejszenie poprzez interpolację, por. np. [23]. W przypadku decymacji konieczne jest automatyczne sprawdzanie, czy dane wynikowe spełniają ograniczenie wynikające z prawa Shannona-Kotielnikowa, por. 3.2.4.

3.3.2. Wstępne przetwarzanie danych

Celem wstępnego przetwarzania jest usunięcie wad występujących w danych surowych lub takie ich przetworzenie, aby można było uzyskać na ich podstawie jak najwięcej informacji przydatnej dla celu identyfikacji. Problem wyboru odpowiednich technik wstępnego przetwarzania danych jest trudny do przedstawienia w postaci uniwersalnych reguł, gdyż wiele zależy tu od intuicji i inżynierskiego wyczucia osoby prowadzącej identyfikację, a także od specyfiki konkretnego zadania. Stąd od systemu wspomagania należy oczekiwać przede wszystkim dostępności typowych usług w tym zakresie, uzupełnionych o pewne ogólne porady dla użytkownika.

Do klasycznych metod wstępnego przetwarzania danych należy filtracja ciągów danych, ich normalizacja, skalowanie, usuwanie składowych deterministycznych, takich jak składowa stała, trend wielomianowy czy składowa okresowa. Omówienie przedstawionych technik przetwarzania danych oraz zasad ich stosowania można znaleźć w wielu podręcznikach, w szczególności w [101]. Usuwanie składowych deterministycznych z ciągów danych może wymagać podania pewnych parametrów dla procedury, takich jak np. stopień trendu czy częstotliwość usuwanej składowej sinusoidalnej, stąd w następnym podpunkcie przedstawiono, w jakim zakresie system wspomagania może wspierać użytkownika w tym względzie.

W przypadku wybrania do pobudzenia obiektu o wielu wejściach sygnału WSWO, por. 3.2.2, możliwe jest wykorzystanie specjalnego sposobu wstępnego przetwarzania danych w celu dekompozycji zadania identyfikacji na obiekty jednowejściowe. Odbywa się to na drodze filtracji w dziedzinie częstotliwości, polegającej na usunięciu z widma sygnału wyjściowego tych częstotliwości, które nie występują w sygnale pobudzającym dla danego wejścia. W ten sposób, stosując odwrotną transformatę Fouriera, można dokonać estymacji niemierzalnego sygnału wyjściowego danego toru sterowania. Metoda ta została opisana w [32] i zostanie tu wykorzystana w identyfikacji przykładowego obiektu elektroakustycznego, por. 3.4.2. W podobny sposób można usunąć z widma sygnału wyjściowego te składowe, które są

szczególnie zakłócone, co pozwala na znaczną poprawę wyników w sytuacji zakłóceń wąskopasmowych, por. 7.2.

3.3.3. Usuwanie składowych deterministycznych

W analizowanych sygnałach mogą występować składowe, które mogą być opisane za pomocą modeli deterministycznych i które należy usunąć z danych przed dalszym przetwarzaniem.

Podstawowym sposobem opisu trendu występującego w ciągu danych jest model wielomianowy, zdefiniowany jako wielomianowa funkcja czasu dyskretnego (reprezentowanego w postaci numeru próbki i):

$$y(i) = \sum_{j=0}^{m} c_j i^j . \tag{3.1}$$

Współczynniki c_j modelu można wyznaczyć przy wykorzystaniu klasycznego kryterium minimalizacji sumy kwadratów różnic między rzeczywistymi (obserwowanymi) wartościami ciągu a wartościami dla modelu wielomianowego, por. np. [110]. Stopień m modelu trendu może być określony przez użytkownika na podstawie wiedzy *a priori*, albo może być wyznaczany automatycznie na drodze iteracyjnej identyfikacji modeli trendu dla kolejnych stopni wielomianu. Wyboru odpowiedniego stopnia wielomianu dokonuje się na drodze analizy odchylenia średniokwadratowego.

Usunięcie trendu wielomianowego z ciągu danych można wykonać odejmując od ciągu danych zidentyfikowany model trendu, co znajduje zastosowanie głównie przy przewidywaniu przyszłych wartości ciągu, albo na drodze odpowiedniego różnicowania ciągu danych, co prowadzi do tzw. modeli sumowalnych, por. [19, 101]. W obu przypadkach dla reszt pozostałych po usunięciu trendu poszukuje się odpowiedniego modelu stochastycznego.

Jeżeli w ciągu danych występują składowe sinusoidalne, to ich usunięcie możliwe jest za pomocą specyficznego sposobu filtracji przez tzw. *wielomiany zerujące*. Filtr o transmitancji

$$1 - 2[\cos(\omega_s \tau_p)]\, z^{-1} + z^{-2}$$

umożliwia wyzerowanie w ciągu danych składowej o postaci $A\sin(\omega_s \tau_p i + \varphi)$. W celu usunięcia składowej sinusoidalnej nie jest więc potrzebna znajomość jej amplitudy A ani fazy φ. Konieczna jest natomiast *dokładna* znajomość częstotliwości ω_s, jako że wynik filtracji jest bardzo silnie zależny od wartości tej częstotliwości.

Jeżeli częstotliwość ta nie jest znana *a priori*, to system wspomagania może pomóc użytkownikowi w jej wyznaczeniu korzystając z nowoczesnych metod wyznaczania widm dyskretnych, takich jak metody *Prony'ego*, *Pisarenki*, *MUSIC* czy *EV*, por. np. [101]. Jednak

osobiste doświadczenie autora w tym względzie pokazuje, że najpewniejszym sposobem wyznaczenia częstotliwości składowych sinusoidalnych jest wykorzystanie klasycznej metody analizy widma sygnału opartej na *periodogramie*, szczególnie jeśli dokona się *zwiększenia gęstości* na osi częstotliwości na drodze uzupełnienia ciągu zerami (*zero-padding*), por. np. [23]. Jeżeli w ciągu występuje więcej składowych sinusoidalnych, to konieczna jest także znajomość ich liczby. W celu określenia liczby składowych sinusoidalnych użytkownik systemu może się posłużyć wykresem *wartości własnych* macierzy autokorelacji odpowiednio dużego wymiaru, por. [67]. Parametry wielomianu zerującego można też próbować wyznaczyć na drodze identyfikacji modelu *AR*, por. [60].

3.4. Planowanie eksperymentu dla identyfikacji obiektu elektroakustycznego

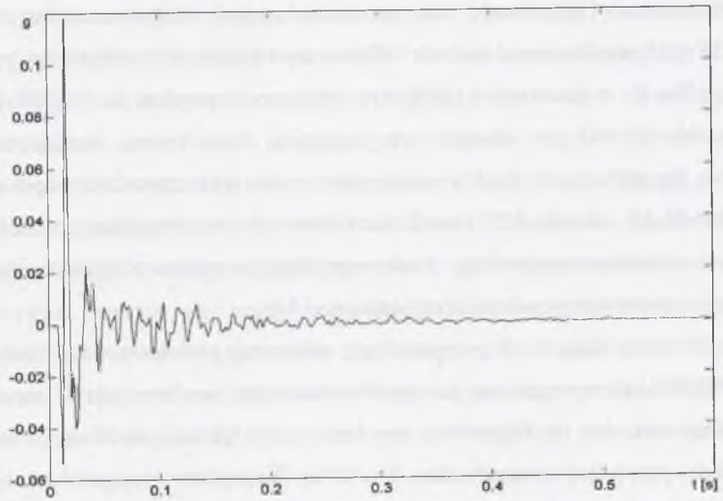
W celu identyfikacji *off-line* przedstawionego w punkcie 2.3 przykładowego obiektu elektroakustycznego należy zaplanować eksperyment, który umożliwi otrzymanie opowiednich modeli torów tego obiektu, potrzebnych do parametryzacji układu ATH. Wymusza to wykorzystanie w eksperymencie identyfikacyjnym tego samego sprzętu, jaki będzie użyty w ATH, a więc sposób wykonywania pomiarów i ich filtracji. Podobnie, częstotliwość próbkowania $f_p = 500$ Hz narzucona jest przez przyjęcie wymagań na zakres tłumionych częstotliwości hałasu. Do określenia pozostaje więc czas trwania eksperymentu i sposób pobudzania obiektu.

Jako kompromis między wymaganiem dokładności modelu, która rośnie wraz z liczbą próbek, a czasem trwania eksperymentu, który powinien być jak najkrótszy ze względu na hałas wytwarzany w trakcie pobudzania obiektu, przyjęto liczbę próbek $N = 4096$. To oznacza, że czas trwania eksperymentu wynosi około 8 s. Wartość ta wynika z poprzednio przeprowadzonych doświadczeń i zapewnia jednocześnie, że czas ten jest znacznie dłuższy niż czas trwania odpowiedzi impulsowej, której przykładowy przebieg dla jednego z torów pokazano na rys. 3.1.

Wybór N jako potęgi liczby 2 stanowi tu naturalną konsekwencję minimalizacji kosztów obliczeń w sytuacji korzystania z szybkiej transformaty Fouriera *FFT*.

Zadanie identyfikacji polega na aproksymacji transmitancji dla wszystkich 12 torów wymienionych w 2.3.2, co powinno być zrobione, zanim układ ATH zostanie uruchomiony. Można to osiągnąć wykonując 3 osobne doświadczenia, gdzie kolejno każdy głośnik zostanie wysterowany specjalnie zaplanowanym ciągiem pobudzającym i zarejestrowane zostaną

sygnały ze wszystkich mikrofonów. Następnie na drodze przetwarzania ciągów pobudzeń i odpowiedzi możliwe będzie zidentyfikowanie odpowiednich transmitancji w sposób wzajemnie niezależny. Podejście takie przedstawiono w [90]. Inną możliwość stanowi jednoczesne pobudzenie wszystkich 3 głośników w tym samym czasie i traktowanie sygnału z mikrofonu jako wyjścia obiektu MISO o 3 wejściach. Pozwala to na skrócenie czasu potrzebnego na przeprowadzenie doświadczeń, ale powoduje, że identyfikacja staje się znacznie trudniejsza. W pracy przedstawione zostanie to drugie podejście.



Rys. 3.1. Typowy przebieg odpowiedzi impulsowej dla obiektu elektroakustycznego
Fig. 3.1. Typical impulse response of the electro-acoustic plant

Aby w trakcie jednego eksperymentu uzyskać maksimum informacji o wszystkich 12 torach, sygnały pobudzające zostały zaprojektowane w specjalny sposób, co omówiono bliżej w następnym podpunkcie. Jednocześnie kilkakrotne powtórzenie eksperymentu umożliwiło potwierdzenie założenia o *powtarzalności*, tj. że dla danej konfiguracji sprzętów w pomieszczeniu system przez pewien czas nie zmienia swoich własności.

3.4.1. Sygnał pobudzający

Podstawowym elementem planowania eksperymentu jest tu wybór sygnałów pobudzających. Typ pobudzenia ma istotny wpływ na wyniki identyfikacji i w literaturze przedmiotu można spotkać wiele propozycji w tym zakresie, por. [39]. Aby spełnić wymaganie *dostatecznego pobudzenia*, por. 3.1, spełnione muszą być dwa podstawowe warunki:

- w celu uzyskania modeli transmitancyjnych o odpowiednio dokładnej charakterystyce amplitudowo-fazowej odpowiednia wartość mocy powinna występować dla wszystkich linii widma do częstotliwości Nyquista;
- sygnał pobudzający na każdym z wejść powinien być nieskorelowany z pozostałymi sygnałami, aby umożliwić identyfikację wszystkich torów na podstawie tylko jednego eksperymentu.

Trzy podstawowe typy sygnałów spełniających powyższe wymagania to pseudolosowy binarny (*PRBS*), pseudolosowy gaussowski oraz wektorowy sygnał wielosinusoidalny ortogonalny (WSWO). W [65] przedstawiono badanie wpływu wymienionych trzech typów sygnałów na wyniki identyfikacji i pokazano, że najlepsze rezultaty można uzyskać dla WSWO. Szczególną zaletą sygnałów WSWO jest własność ortogonalności, dzięki czemu stosując odpowiednie przetwarzanie sygnałów w dziedzinie częstotliwości, można wykorzystać metodę dekompozycji zadania identyfikacji modelu *MISO* na 3 niezależne zadania identyfikacji modeli *SISO*, co omówiono w następnym podpunkcie, a także poprawić wyniki w przypadku wykonywania doświadczeń w obecności sygnału zakłócającego, por. 7.2.

Ponieważ liczba danych N powinna być całkowitą wielokrotnością liczby sinusoid w sygnale WSWO, więc przyjmując, że sygnał pobudzający powinien składać się z $K = 2048$ sinusoid, zapewnia się jednocześnie uzyskanie największej możliwej rozdzielczości charakterystyk częstotliwościowych dla $N = 4096$. Wymaganie ortogonalności najłatwiej spełnić przydzielając kolejne harmoniczne częstotliwości podstawowej $\Omega = 2\pi / K$ do kolejnego wejścia k, $k = 1,2,3$:

$$u_k(i) = \sum_{j=0}^{\frac{K-2}{3}} A \sin\left[(3j + k - 1)\Omega i + \Phi_j\right], \tag{3.2}$$

gdzie fazy Φ_j można generować w sposób losowy, albo zgodnie z algorytmem minimalizującym współczynnik szczytu, por. 3.2.2.

Amplitudę A dobrano doświadczalnie tak, aby wszystkie urządzenia występujące w obiekcie pracowały w zakresie liniowej części charakterystyki statycznej. W ten sposób uzyskano spełnienie dwóch wymagań: poszczególne sygnały wejściowe są wzajemnie ortogonalne oraz spełniony jest warunek nieustającego pobudzenia, gdyż wszystkie tory są pobudzane w pełnym zakresie częstotliwości i amplitud, jakie mogą się pojawić w trakcie stosowania.

3.4.2. Wstępne przetwarzanie danych

Zazwyczaj dane zebrane w trakcie eksperymentu identyfikacyjnego podlegają pewnym procedurom wstępnego przetwarzania, zanim mogą zostać wykorzystane w procedurze identyfikacji. Tutaj podstawową czynność stanowi usuwanie wartości stałej (*offsetu*) z sygnałów zebranych z mikrofonów (wyjść obiektu). W tym celu, przed rozpoczęciem pobudzania obiektu, mierzono wartości sygnałów przez czas około 100 ms i obliczoną wartość średnią traktowano jako ocenę *offsetu*, którą następnie odejmowano od danych zebranych w trakcie eksperymentu.

Wykorzystując własność ortogonalności sygnałów typu WSWO można także dokonać dekompozycji problemu identyfikacji modelu *MISO* na niezależne zadania identyfikacji modeli *SISO* dla każdego toru sterowania z osobna, por. [32]. Odbywa się to na drodze specjalnej filtracji w dziedzinie częstotliwości, polegającej na usunięciu z widma sygnału wyjściowego wszystkich tych składowych, które nie występują w widmie danego sygnału wejściowego. W ten sposób, po zastosowaniu odwrotnej transformaty Fouriera dla tak przefiltrowanego sygnału można uzyskać ocenę przebiegów czasowych *wyjść cząstkowych* dla poszczególnych torów sterowania. Przetwarzając następnie kolejno poszczególne pary odpowiadających sobie sygnałów wejścia i estymowanego *wyjścia cząstkowego*, można uzyskać model danego toru przy znacznie zmniejszonym nakładzie obliczeniowym oraz zredukowanym wpływie pozostałych wejść. W punkcie 7.1 porównano wyniki identyfikacji obiektów *MISO* uzyskane bez filtracji umożliwiającej dekompozycję oraz z zastosowaniem dekompozycji. Dodatkowo, w sytuacji gdy w trakcie eksperymentu występuje sygnał zakłócający o widmie dyskretnym lub wąskopasmowym, to stosując filtrację w dziedzinie częstotliwości polegającą na usunięciu składowych widma odpowiadających częstotliwościom zakłócenia, można znacząco poprawić wyniki identyfikacji, por. 7.2.

Rozdział 4

Struktura modelu

Zakładając spełnienie omówionych w rozdziale 3 warunków poprawnego wykonania eksperymentu, a w szczególności, że prawdziwe jest założenie o rozdzielności obiektu **S** od źródła pobudzenia **R**, przyjmuje się w dalszej części, że celem identyfikacji jest znalezienie modelu dla **S**.

Niech *model M* oznacza pewien system, którego wyjście *y* przyjmuje wartości w tej samej przestrzeni co wyjście **y** obiektu **S** w odpowiedzi na to samo pobudzenie **u**. Aby móc wyznaczyć model *M*, trzeba najpierw określić podprzestrzeń, w której *M* się znajduje. Cała sztuka polega tu na określeniu tej przestrzeni jako innej (znacznie mniejszej) niż przestrzeń, w której występuje **S**, co jest zgodne z intuicyjnie oczywistym warunkiem, że za pomocą równań matematycznych nie można opisać dokładnie procesu fizycznego oraz z tym, że dane doświadczenie może dostarczyć jedynie ograniczonej informacji. Ograniczenie przestrzeni odbywa się na drodze przyjęcia pewnej *struktury* $\mathcal{M}$, definiowanej jako zbiór wszystkich „próbnych" modeli *M* posiadających tę samą postać [17].

Podstawowym ograniczeniem dla struktury modelu jest wymaganie *identyfikowalności*, por. np. [17, 79, 96]. Problem identyfikowalności jest złożony i w literaturze przedmiotu można spotkać się z różnym podejściem do tego zagadnienia. Wyróżnia się np. identyfikowalność systemową, związaną z możliwością wyznaczenia współczynników modelu dla danej struktury, oraz identyfikowalność parametryczną, odpowiadającą na pytanie, czy możliwe jest wyznaczenie parametrów fizycznych, gdy znane są współczynniki modelu wejścia-wyjścia. Dla celów tej pracy interesujący jest ten pierwszy rodzaj identyfikowalności. Zasadniczo pojęcie identyfikowalności związane jest z warunkami uzyskania jednoznacznego rozwiązania zadania identyfikacji. Warto tu zauważyć, że o ile takie pojęcia, jak sterowalność czy obserwowalność, wynikają bezpośrednio z własności opisywanego systemu, to identyfikowalność zależy także od tego, *jak* przeprowadzone zostało doświadczenie identyfikacyjne [96].

W celu uzyskania opisu obiektu w postaci oszczędnej należy model *M sparametryzować* w odpowiedni sposób. Opis ten z jednej strony musi uwzględniać cel, dla jakiego jest tworzony, a z drugiej własności identyfikowanego obiektu. W punkcie 4.1 omówiono zadanie modelowania polegające na przyjęciu struktury $\mathcal{M}$ dla modelu stochastycznego. Ze względu na

zalety powszechnie stosowanego opisu w postaci wejściowo-wyjściowej oraz modelowania liniowego w dalszej części rozdziału skupiono się na tej klasie modeli.

Powszechnie stosowane są także modele *nieparametryczne*, gdy niewiadome są funkcjami rzeczywistymi lub ciągami nieskończonymi, jak np. *gęstość widmowa mocy* czy *odpowiedź impulsowa*. Jednak wielkości, które można wyznaczyć na podstawie danych, muszą stanowić *skończenie wymiarowe aproksymacje* tych nieskończenie wymiarowych funkcji teoretycznych. Na przykład, jeśli gęstość widmowa mocy ma być aproksymowana za pomocą periodogramu, to liczba wyznaczonych prążków powinna być mniejsza niż liczba danych, gdyż w przeciwnym przypadku odchylenie standardowe periodogramu równe będzie jego wartościom, por. np. [11]. Podobnie w przypadku nieparametrycznego modelu odpowiedzi impulsowej, jeśli obiekt jest stabilny, a jednocześnie zakłócany, to liczba współczynników tego modelu powinna być obcięta, co prowadzi do parametrycznego modelu skończonej odpowiedzi impulsowej *FIR*. Stąd w przypadku modeli nieparametrycznych można liczbę punktów, dla których funkcja jest wyznaczana, uważać także za swego rodzaju strukturę modelu. Jednak ponieważ najczęściej do modelowania i analizy różnego rodzaju procesów wykorzystuje się modele stochastyczne przedstawione w postaci *parametrycznej*, stąd w dalszej części rozpatrywane są struktury tego typu modeli.

4.1. Taksonomia modeli parametrycznych

Celem modelowania, polegającego na przyjęciu struktury $\mathcal{M}$, jest *ograniczenie* zbioru wszystkich możliwych modeli w taki sposób, aby liczba stopni swobody w tym zbiorze była odpowiednio mała, tj. by w trakcie identyfikacji możliwe było wybranie odpowiedniego modelu z tego zbioru na podstawie dostępnych danych. Trudność polega na tym, aby przy okazji takiej redukcji nie odrzucić przypadkiem tych struktur, które byłyby *odpowiednie* do postawionego celu.

4.1.1. Modele stochastyczne

Informacja o strukturze obiektu **S** może być określona w postaci *modelu probabilistycznego* $p(y \mid u, M)$. Taki zapis pozwala zaznaczyć przyczynową zależność (lub niezależność) poprzez wprowadzenie lub usunięcie zmiennych po prawej stronie znaku warunkowości |. W praktyce oczywiście trudno by było wykorzystać tak ogólną postać bezpośrednio, stąd potrzebne są odpowiednie sposoby opisu modeli probabilistycznych przy użyciu struktur, które są mniej ogólne, a więc zawierają więcej informacji *a priori*.

Ponieważ nie da się wyznaczyć wartości dla większej liczby niewiadomych niż (skończona) liczba zebranych danych, więc model powinien zostać w jakiś sposób *sparametryzowany*. Podstawowym i powszechnie stosowanym sposobem określenia rozkładu prawdopodobieństwa dla y jest wykorzystanie tzw. *modeli stochastycznych*, które umożliwiają generację wektora losowego y na podstawie innego wektora losowego e o zadanym rozkładzie $p(e)$ za pomocą pewnego algorytmu $y = F(u, e)$. Jeśli $p(e)$ jest znane, np. składowe e są niezależnymi zmiennymi gaussowskimi o zerowej średniej i jednostkowej wariancji, to F determinuje rozkład y. W [17] wykazano, że dla e o rozkładzie normalnym zawsze można znaleźć takie F, które umożliwi otrzymanie dowolnego rozkładu $p(y)$.

W modelu stochastycznym można wyróżnić postać $\mathcal{F}$ (np. model w postaci równania różnicowego), zbiór wskaźników struktury μ (parametry całkowitoliczbowe określające wymiar, np. stopień równania) oraz wektor parametrów modelu θ (zbiór współczynników rzeczywistych), co można zapisać w ogólnej postaci jako

$$y = \mathcal{F}(u, e, \mu, \theta). \tag{4.1}$$

$\mathcal{F}$ (czyli postać równania) powinno być znane *a priori*, μ może być znane lub nie, natomiast nieznane są e i θ. Znany powinien być także wymiar n zbioru μ, gdyż zależy on od $\mathcal{F}$, natomiast wymiar m wektora θ zależy od nieznanego μ i powinien być znacznie mniejszy niż wymiar zbioru danych $(\mathbf{U}_N, \mathbf{Y}_N)$.

Przy takiej parametryzacji model stochastyczny można zapisać jako trójkę $M = (\mathcal{F}, \mu, \theta)$, przy czym jego *struktura* określona jest przez $\mathcal{M} = (\mathcal{F}, \mu)$. Identyfikacja modelu dynamicznego sprowadza się więc do wyznaczenia wektora parametrów θ dla założonej struktury $\mathcal{M}$ na drodze minimalizacji funkcji kosztów

$$V(\theta \mid \mathcal{M}, \mathbf{U}_N, \mathbf{Y}_N) = -\ln L(\theta \mid \mathcal{M}, \mathbf{U}_N, \mathbf{Y}_N)/N \tag{4.2}$$

W modelu stochastycznym występuje oddzielenie dwóch różnych cech procesu: pewnych *stałych własności*, które mogą być wykorzystane przy wszelkiego rodzaju projektowaniu bazującym na modelu tego procesu, od *losowości*, która wprawdzie wpływa na odpowiedź tego procesu, ale której nie uwzględnia się na etapie stosowania. Trzy niewiadome we wzorze (4.1) pełnią różne role. Zbiory μ i θ niosą informację o procesie, która jest niezmienna w czasie i stąd może być użyteczna w fazie stosowania modelu, natomiast zmienna elementarna e jest losowa. Informacja o niej jest prawdziwa tylko dla konkretnych danych $(\mathbf{u}, \mathbf{y})$. A to oznacza, że poprzez e uwzględnia się to, że ten sam eksperyment, przeprowadzony na tym samym obiekcie, może generować różne dane $(\mathbf{u}, \mathbf{y})$. Jeśli eksperyment zostanie powtórzony kiedy indziej, to powinny pozostać niezmienione wartości μ i θ, ale nie wartości e.

Ponieważ zadanie identyfikacji modelu stochastycznego może obejmować nie tylko wyznaczenie zbioru liczb rzeczywistych θ (parametrów modelu), ale także zbioru μ, więc warto jest w tym celu zdefiniować *ciąg struktur hierarchicznych* $\{\mathcal{M}(m)\}$. Szeregowanie takie można zrealizować np. w ten sposób, by $\mathcal{M}(m-1) \subset \mathcal{M}(m)$, $\mathcal{M}(0) = \phi$, gdzie m określa liczbę parametrów wymaganą do opisu modelu dla zbioru wskaźników struktury μ_m. Jeśli μ jest skalarem, to szeregowanie takie jest naturalne, jeśli μ jest wektorem (co zachodzi w większości przypadków), to można to zrobić na wiele różnych sposobów. Liczbę całkowitą m można uważać za *miarę złożoności* modelu. Warto jednak pamiętać, że mamy tu do czynienia z porządkiem częściowym i nie zawsze można uporządkować dwie struktury.

Rozpatrywanie hierarchii struktur jest przydatne w procedurze poszukiwania właściwej struktury modelu, por. 5.3.2. Zgodnie z fundamentalną zasadą prostoty, znaną jako tzw. „brzytwa Ockhama" (łac. *Entia non sunt multiplicanda necessitatem*), należy wybrać najprostszą strukturę, która spełni zadane kryterium wyboru, a więc szeregowanie struktur w innym porządku niż w kierunku wzrostu złożoności nie ma sensu. Ponadto niektóre metody testowania statystycznego stosowane do falsyfikacji struktury modelu $\mathcal{M}(m)$ jako zbyt prostej, aby umożliwić uzyskanie modelu niesprzecznego z danymi, wymagają określenia struktury alternatywnej, bogatszej niż struktura porównywana (np. *test F*, por. 5.1.2). Warunek ten można nazwać *zagnieżdżaniem* (ang. *nesting*), por. [17]. Oznacza on, że każdemu zbiorowi wskaźników struktury modelu $\mu_{m'}$ odpowiada zbiór wskaźników $\mu_{m''}$ taki, że $m'' > m'$. Zagnieżdżanie może być osiągnięte w różny sposób, np. strukturę uboższą można uzyskać z bogatszej przez wyzerowanie części parametrów. Przykładem *prostej struktury zagnieżdżonej* może być wymierna funkcja przejścia, gdzie μ jest wspólnym rzędem licznika i mianownika, a θ zawiera $m = 2\mu + 1$ współczynników. Funkcja przejścia niższego rzędu może być więc uważana za specjalny przypadek funkcji wyższego rzędu, w której niektóre współczynniki mają wartość zero. Struktura jest uważana za *złożoną*, jeśli dla danego m można podać więcej niż jeden zbiór wskaźników μ_m, np. dla $m = 2$ można podać trzy dwójki: (0,2), (1,1) i (2,0).

4.1.2. Modele wejściowo-wyjściowe w postaci czarnej skrzynki

Funkcja wiarygodności dla modelu dynamicznego wynika z (2.9):

$$L(M \mid \mathbf{U}_N, \mathbf{Y}_N) = \prod_{i=1}^{N} p[\mathbf{y}(i) \mid \mathbf{U}_{i-1}, \mathbf{Y}_{i-1}, M]. \tag{4.3}$$

W zależności od sposobu obliczania $p[\mathbf{y}(i) \mid \mathbf{U}_{i-1}, \mathbf{Y}_{i-1}, M]$ można wyróżnić tzw. modele *wejściowo-wyjściowe* lub w postaci *równań stanu*.

Model *wejściowo-wyjściowy* opisuje zależności *tylko* między *obserwowanymi* zmiennymi u i y, a więc w przypadku obiektu dynamicznego jego *wejściowo-wyjściowy* model stochastyczny można zapisać w postaci

$$y(i) = F[U_{i-1}, Y_{i-1}, e(i) \mid M] \tag{4.4}$$

Dla modelu tego funkcję wiarygodności można obliczyć, przy założeniu że jest on *odwracalny*, tzn. da się wyznaczyć e za pomocą algorytmu odwrotnego

$$e(i) = F^{-1}[U_{i-1}, Y_i, \mid M]. \tag{4.5}$$

Wykorzystując model odwrotny (4.5) oraz dane $(\mathbf{U}_{i-1}, \mathbf{Y}_i)$, można wyznaczyć realizacje zmiennej e w postaci tzw. ciągu *innowacji* zgodnie z zależnością:

$$\mathbf{e}(i \mid M) = F^{-1}[\mathbf{U}_{i-1}, \mathbf{Y}_i, \mid M]. \tag{4.6}$$

Innowację $\mathbf{e}(i)$ można interpretować jako część $\mathbf{y}(i)$ zawierającą informację nową, której nie ma w $(\mathbf{U}_{i-1}, \mathbf{Y}_{i-1})$ [4].

Model stochastyczny nazywany jest także *modelem predykcyjnym*, ponieważ może być wykorzystany do predykcji wyjścia $y(i)$ na podstawie danych (U_{i-1}, Y_{i-1}):

$$y(i \mid i-1) = F[U_{i-1}, Y_{i-1} \mid M]. \tag{4.7}$$

Wiele metod identyfikacji modeli wejściowo-wyjściowych bazuje na minimalizacji *błędu predykcji jednokrokowej* $e(i \mid i-1) = y(i) - y(i \mid i-1)$, por. np. [79, 132].

W odróżnieniu od modelu wejściowo-wyjściowego, model w postaci *równań stanu* opisuje zależności także między innymi zmiennymi niż tylko te obserwowane. W przypadku modelu stochastycznego w postaci równań stanu

$$x(i) = F^x[U_{i-1}, X_{i-1}, e^x(i) \mid M^x], \tag{4.8a}$$

$$y(i) = F^y[U_{i-1}, X_i, e^y(i) \mid M^y] \tag{4.8b}$$

zmienne losowe elementarne występują zarówno w *równaniach wyjścia* (4.8b), jak i w *równaniach stanu* (4.8a) – jest ich więc *więcej*, niż wynosi liczba składowych y. Ponieważ jednocześnie więcej detali jest tu wbudowanych w model, więcej tych detali musi być także znanych *a priori*, jako że z pozyskanych danych można uzyskać tylko pewną ilość informacji.

Tak więc w porównaniu do modelu wejściowo-wyjściowego, model równań stanu wymaga znacznie większej wiedzy *a priori* o systemie, a także wprowadzenia dodatkowych zmiennych losowych. Co więcej, trzeba założyć, że obie zmienne e^x i e^y są statystycznie *niezależne*. Podejście takie jest uzasadnione w sytuacji wykorzystania *filtru Kalmana* jako estymatora stanu, por. np. [54, 4], ale to nie jest zadanie identyfikacji, bo parametry obiektu są tu znane, poszukiwane są tylko niemierzalne współrzędne stanu procesu x. W przypadku identyfikacji,

tzn. estymacji zarówno współrzędnych stanu, jak i parametrów modelu, wymagane jest stosowanie *rozszerzonego filtru Kalmana*, por. [4, 78]. Jednak to zadanie jest już znacznie bardziej skomplikowane, a nie zawsze konieczne do wykonania, np. dla celów predykcji czy sterowania ze sprzężeniem zwrotnym model w postaci opisu wewnętrznego nie jest potrzebny.

Ponieważ omówione wyżej typy modeli wymagają wprowadzenia różnej ilości informacji *a priori*, więc można je powiązać także z innym podziałem, w którym rozróżnia się tzw. modele *czarnej skrzynki* (ang. *black-box*), które nie korzystają z żadnej informacji *a priori*, *szarej skrzynki* (*grey-box*), korzystające z niej częściowo, oraz *białej skrzynki* (*white-box*), które bazują wyłącznie na takiej informacji [128]. Różnica między nimi polega na tym, ile *znanych a priori* szczegółów jest zawartych w poszczególnej formie, a ile pozostało do identyfikacji - im skrzynka mniej *biała*, tym więcej informacji trzeba *uzyskać z danych*. Kompromis zależy od tego, w jakim stopniu można rzeczywiście polegać na każdym ze źródeł informacji, tj. na *wiedzy a priori* oraz *danych*, a także od tego, ile informacji jest wymagane.

Modele typu *czarna skrzynka* mają bliski związek z modelami *wejściowo-wyjściowymi* i stosowaną do ich identyfikacji metodą *ML*. Natomiast w modelach typu *szara skrzynka* informację *a priori* wprowadza się albo poprzez prawdopodobieństwo *a priori* i użycie podejścia bayesowskiego, tj. metody *MAP*, albo wykorzystując opis w postaci równań stanu, umożliwiający wprowadzenie wiedzy strukturalnej o fizycznej naturze procesu. Algorytmy wykorzystujące obie możliwości przedstawiono np. w [70] i [140].

Podsumowując, podstawowe zalety modeli typu czarna skrzynka, takie jak niewielkie wymagania na wiedzę *a priori* oraz mniejsze ograniczenia na strukturę modelu $\mathcal{M}$, a jednocześnie fakt, że taki sposób modelowania jest wystarczający dla większości spotykanych w praktyce celów identyfikacji, sprawiają, że najczęściej identyfikowane są właśnie modele czarnej skrzynki. Ponadto, skoro obiektywna informacja o procesie reprezentowana jest w postaci pomiarów jego wejść i wyjść, to model wejściowo-wyjściowy stanowi jednocześnie naturalny sposób opisu procesu. Dlatego też w omawianym w pracy zagadnieniu komputerowego wspomagania identyfikacji główna uwaga została poświęcona identyfikacji właśnie takich modeli.

4.1.3. Modele liniowe i nieliniowe

W dotychczasowych rozważaniach problem nieliniowości nie był rozpatrywany w sposób jawny, był po prostu ukryty w funkcji wiarygodności $L(\theta \mid \mathcal{M}, \mathbf{U}_N, \mathbf{Y}_N)$. Jednak metody identyfikacji modeli nieliniowych mogą sprawiać wiele problemów, np. jeśli model przejawia jakieś zachowanie chaotyczne dla pewnych wartości parametrów θ, to $L(\theta \mid \mathcal{M}, \mathbf{U}_N, \mathbf{Y}_N)$ będzie

bardzo złożoną funkcją tych parametrów, powodując, że występujące w procedurze identyfikacji zadania aproksymacji, walidacji oraz falsyfikacji, por. 5.1, mogą być bardzo trudne lub wręcz niemożliwe do rozwiązania.

Do popularności modeli w postaci liniowej przyczynia się również to, że modelowanie nieliniowe wymaga na ogół sporych umiejętności. Podejście liniowe jest często wystarczające, jeżeli stosowanie modelu wymaga jego poprawności tylko w sąsiedztwie pewnej trajektorii nominalnej; w szczególności dotyczy to przypadku sterowania obiektem wokół tej trajektorii. Procesy nieliniowe aproksymuje się często za pomocą modeli liniowych dla niewielkich *odchyłek* sygnału wejściowego, a jeśli sterowanie procesu przebiega właściwie, to jego zachowanie nie odbiega zbytnio od założonej trajektorii nominalnej. W takim przypadku model liniowy stanowi wystarczająco dokładny opis dla danego celu. Kluczowym punktem jest tu *możliwość linearyzacji*, tj. możliwość rozwinięcia w szereg Taylora wokół trajektorii nominalnej. To wymaganie wyklucza wszelkie modelowanie zachowań zawierających nieciągłości, takie jak np. tarcie suche, histereza, luzy itp. Konieczne jest także, aby model liniowy był *stabilny* wokół trajektorii – małe odchyłki dla warunku początkowego oraz sygnału pobudzającego nie mogą powodować zbyt dużych różnic w odpowiedziach.

Przy modelowaniu liniowym występuje naturalna tendencja do opisu w postaci *czarnej skrzynki* nawet wtedy, gdy dostępna jest informacja *a priori*, np. na temat zakresu stałych czasowych. Modelowanie nieliniowe także często opiera się tylko na niewielkiej ilości informacji *a priori*. Najczęściej stosowanymi nieliniowymi modelami typu *czarna skrzynka* są modele w postaci szeregów Volterry, por. np. [121], modele blokowe złożone z liniowych bloków dynamicznych oraz nieliniowych statycznych, jak modele Hammersteina i Wienera, por. np. [5, 6, 14, 42, 44], oraz modele biliniowe, por. np. [91, 137] - z tej prostej przyczyny, że właśnie dla takich modeli zaproponowano metody estymacji ich parametrów. Oprócz podejścia parametrycznego, coraz popularniejsze jest także tworzenie opisu nieliniowego z wykorzystaniem sieci neuronowych (ang. *neural networks*), zbiorów rozmytych (*fuzzy sets*) czy falek (*wavelets*), por. np. [94]. Przegląd różnych struktur oraz metod pod kątem jednolitego spojrzenia na identyfikację modeli nieliniowych typu *czarna skrzynka* przedstawiono w [128], natomiast podstawy matematyczne stosowanych aproksymacji nieliniowych zawiera [52].

Jednak dopasowanie nieliniowego modelu typu *czarna skrzynka* jest na ogół zadaniem bardzo trudnym i wymaga wykonania dużo więcej obliczeń niż w przypadku liniowym. Identyfikacja *struktury* modelu jest znacznie trudniejsza, a czasem wręcz niemożliwa ze względu na bardzo dużą liczbę możliwych kombinacji nieliniowych elementów w ramach ogólnej klasy modeli nieliniowych.

Podsumowując, mimo że przy bardzo dokładnej analizie własności procesów rzeczywistych często można zauważyć, że są one nieliniowe, to jednak zazwyczaj dąży się do tworzenia ich opisu matematycznego w postaci liniowej, ponieważ:

- metody analizy dynamicznych modeli liniowych są zarówno efektywne, jak i stosunkowo proste w użyciu oraz umożliwiają łatwe przejście z jednej formy opisu na inną (np. z równania różnicowego na postać transmitancyjną);
- własności modeli liniowych są intuicyjnie zrozumiałe (np. odpowiedź skokowa, charakterystyka amplitudowo-fazowa);
- wiele zastosowań wymaga dostarczenia modelu w postaci liniowej; szczególnie dotyczy to sterowania i predykcji;
- modele nieliniowe są znacznie trudniejsze do identyfikacji, brak jest spójnych, dobrze przetestowanych i powszechnie stosowanych metod identyfikacji [107];
- modele nieliniowe mogą przejawiać różne nieoczekiwane zachowania, czasem wręcz o charakterze *chaosu*, które wcale nie muszą przypominać normalnego zachowania procesu, por. np. [124].

Z przedstawionych wyżej względów, a także z uwagi na to, że w zastosowaniach dominują jednak modele liniowe, w dalszej części pracy zaproponowano szczegółowe rozwiązanie zadania komputerowego wspomagania identyfikacji dla klasy *modeli liniowych*.

4.2. Liniowy model wejściowo-wyjściowy

W zależności od tego, czy przebiegi czasowe wejść i wyjść procesu można opisać skalarnymi czy wektorowymi funkcjami czasu, wyróżnia się modele *SISO*, *MISO* oraz *MIMO*. Jeżeli brak jest pomiarowo dostępnych wejść, to proces opisuje się modelem ciągu czasowego (wyjściowego): skalarnym lub wektorowym. Modele o jednym wyjściu uważa się za modele skalarne, bez względu na liczbę wejść, co wynika z podobieństwa metod estymacji parametrów dla tych modeli. Natomiast procesy o wielu wyjściach mogą być identyfikowane albo w postaci zbioru modeli skalarnych, osobno dla każdego elementu wektorowego ciągu wyjściowego, albo łącznie jako jeden model wektorowy. Przy tym pierwszym podejściu identyfikacja jest prostsza do wykonania, ale nie umożliwia wyznaczenia bezpośredniej zależności między elementami ciągu wyjściowego. W drugim przypadku identyfikacja jest trudniejsza, przede wszystkim ze względu na występującą możliwość pojawienia się niejednoznaczności w określeniu struktury

modelu, por. np. [107], chociaż możliwe jest przedstawienie całego modelu w postaci bardziej zwartej oraz określenie zależności między elementami ciągu wyjściowego.

Z uwagi więc na to, że nawet w przypadku obiektu o wielu wyjściach opis w postaci zbioru modeli skalarnych jest wygodniejszy, gdyż w naturalny sposób umożliwia przedstawienie go w postaci transmitancyjnej, w przeciwieństwie do stosowanego dla modeli wektorowych opisu w postaci *MFD* (*Matrix Fraction Description*) [53], w dalszym ciągu rozpatrywane będą modele skalarne. Ponieważ jednocześnie modele *SISO* lub modele skalarnych ciągów czasowych można uważać za szczególny przypadek modeli *MISO*, stąd dalsze rozważania ograniczone zostaną do modeli *MISO*.

Przebieg sygnału wyjściowego jest efektem nie tylko pobudzenia procesu za pomocą odpowiedniego przebiegu sygnału wejściowego, ale także oddziaływania otoczenia, por. 2.2.3, więc wyjście y obiektu **S** jest traktowane jako suma dwóch składowych: wyjścia *toru sterowania* oraz hipotetycznego *zakłócenia* sprowadzonego na wyjście obiektu. Zakłócenie to odzwierciedla sumaryczny efekt oddziaływania na wyjście wszelkich innych zmiennych poza wejściami. Innymi słowy, przedstawia przebieg czasowy wyjścia procesu, na którego wejściu jest stale zero (oczywiście po rozładowaniu warunków początkowych).

Najwygodniejszym sposobem modelowania niepewności jest wykorzystanie równania stochastycznego przedstawionego w sposób ogólny w postaci (4.4). Przyjmując, że w obiekcie występuje q wejść u_k, $k = 1, 2, \ldots, q$, oraz zakładając, że zależności wiążące wejścia i wyjście są *liniowe* i *niezmienne w czasie*, można model obiektu przedstawić w postaci:

$$y(i) = \sum_{k=1}^{q} G_k(z^{-1})\, u_k(i) + H(z^{-1})\, e(i), \tag{4.9}$$

gdzie $G_k(z^{-1})$ jest *transmitancją k-tego toru sterowania*, a sumaryczne zakłócenie sprowadzone na wyjście obiektu jest modelowane jako sygnał losowy powstały na drodze filtracji białego szumu $e(i)$ przez filtr liniowy $H(z^{-1})$, por. np. [79, 101].

Zadanie identyfikacji modelu (4.9) polega na wyznaczeniu wektora θ zawierającego współczynniki transmitancji dla założonej struktury, zgodnie z przyjętym kryterium aproksymacji. Według zasady *ML* estymacja parametrów θ odbywa się na drodze minimalizacji funkcji strat, zgodnie z (4.2), co przy założeniu że $e(i)$ jest białym szumem o rozkładzie normalnym, sprowadza się do minimalizacji sumy kwadratów błędów predykcji [79]:

$$\hat{\theta} = \arg\min \frac{1}{N} \sum_{i=1}^{N} \varepsilon^2(i \mid i-1), \tag{4.10}$$

gdzie błąd predykcji wyznacza się z równania:

$$\varepsilon(i \mid i-1) = H^{-1}(z^{-1})\,[\,y(i) - \sum_{k=1}^{q} G_k(z^{-1})\, u_k(i)]. \tag{4.11}$$

W zależności od sposobu parametryzacji transmitancji $G_k(z^{-1})$ i $H(z^{-1})$ wyróżnia się różne struktury modeli, por. 4.2.1, oraz stosuje różne algorytmy minimalizacji funkcji strat w (4.10), por. dodatek A.

Zamiast definiować funkcję strat w dziedzinie czasu, można także wyznaczyć wektor parametrów θ na drodze minimalizacji funkcji strat V zdefiniowanej w dziedzinie częstotliwości po wykonaniu dyskretnej transformacji Fouriera na ciągach danych, por. np. [89]. Podejście to jednak nie jest popularne ze względu na większe problemy obliczeniowe, niż występujące w zadaniu (4.10), oraz trudności w identyfikacji modeli wyższych rzędów.

Wadą modeli w postaci transmitancji dyskretnych w czasie występujących w (4.9) jest to, że poszczególne parametry w modelu nie mają interpretacji fizycznej, stąd czasami użytkownik oprogramowania do identyfikacji może oczekiwać także przedstawienia jakiejś aproksymacji za pomocą prostego modelu w postaci transmitancji ciągłej w czasie, w której można by wyróżnić pewne parametry określające własności dynamiczne modelu, takie jak czas opóźnienia, stała czasowa, częstotliwość drgań własnych, współczynnik tłumienia, współczynnik wzmocnienia itp. Parametry te mogą być przydatne np. przy doborze parametrów regulatora PID. W związku z tym w przedstawionym w rozdziale 6 systemie *MULTI-EDIP* zaproponowano także możliwość aproksymacji modeli torów sterowania $G_k(z^{-1})$ za pomocą transmitancji operatorowych ciągłych $\hat{G}_k(s)$ standardowych elementów dynamicznych pierwszego lub drugiego stopnia (inercyjnych lub oscylacyjnych), minimalno- lub nieminimalnofazowych. Aproksymacja ta polega na oszacowaniu parametrów $\hat{G}_k(s)$ dla wybranych kilku standardowych struktur modelu ciągłego w taki sposób, aby zminimalizować funkcję strat definiowaną jako suma kwadratów różnicy odpowiedzi skokowej otrzymanego modelu dyskretnego i spróbkowanej odpowiedzi modelu ciągłego. Ponieważ funkcja strat jest silnie nieliniową funkcją parametrów, stąd do jej minimalizacji zastosowano zmodyfikowaną metodę Newtona-Raphsona, por. [51]. Wybór struktury modelu ciągłego dokonywany jest automatycznie na podstawie minimum funkcji strat. Podkreśla się, że przedstawiony wyżej sposób aproksymacji za pomocą modelu ciągłego różni się od typowego podejścia do identyfikacji modeli z czasem ciągłym na podstawie dyskretnych pomiarów, por. np. [93, 49].

4.2.1. Typ i struktura liniowego modelu MISO

Transmitancje dyskretne $G_k(z^{-1})$ oraz $H(z^{-1})$ wygodnie jest sparametryzować za pomocą odpowiednich wielomianów operatora z^{-1}. W zależności od założeń odnośnie do własności dynamicznych w poszczególnych torach oraz sposobu parametryzacji transmitancji można wyróżnić różne struktury. Szczegóły dotyczące parametryzacji poszczególnych struktur liniowych modeli stochastycznych oraz metod estymacji ich współczynników przedstawiono w [101], gdzie jednocześnie wprowadzono pojęcie *typu* modelu liniowego, definiowanego odpowiednio do sposobu parametryzacji transmitancji $G_k(z^{-1})$ i $H(z^{-1})$.

Wprowadzenie pojęcia *typ modelu* jest wygodne w praktyce, gdyż o ile zdefiniowanie w punkcie 4.1.1 struktury modelu $\mathcal{M}$ jako dwójki $(\mathcal{F}, \mu)$, gdzie $\mathcal{F}$ oznacza postać modelu, a μ wektor liczb całkowitych (zbiór wskaźników) zależny od $\mathcal{F}$, było przydatne dla celów dyskusji ogólnych problemów identyfikacji, to jednak dla użytkownika systemu komputerowego wspomagania identyfikacji definicja taka jest zbyt ogólna. Wynika to zarówno z potrzeb użytkownika, który chce dopasować cel identyfikacji oraz wiedzę *a priori* do zaprogramowanych metod, jak i ze względów konstrukcyjnych, ponieważ z różnymi postaciami modelu związane są różne metody identyfikacji czy testowania, a także różny sposób prezentacji modelu.

Aby jednak nie komplikować zapisu, nie wprowadza się tu dodatkowego oznaczenia na *typ* modelu, lecz jedynie przyjmuje się, że postać $\mathcal{F}$ liniowego modelu stochastycznego określona jest przez jego *typ* wynikający ze sposobu parametryzacji $G_k(z^{-1})$ oraz $H(z^{-1})$. Wektor liczb całkowitych μ, będący zbiorem wskaźników strukturalnych zawierającym stopnie poszczególnych wielomianów w modelu (a także wartości czasów opóźnienia w poszczególnych torach sterowania), nazwany będzie dalej *strukturą typu*. Natomiast występujący w ogólnej postaci modelu $M = (\mathcal{F}, \mu, \theta)$ wektor parametrów θ zawiera współczynniki wielomianów operatora z^{-1}, za pomocą których sparametryzowano transmitancje $G_k(z^{-1})$ oraz $H(z^{-1})$.

W zależności od sposobu tej parametryzacji wyróżnia się następujące typy modeli [101]:

- *ARX*

$$y(i) = \sum_{k=1}^{q} z^{-d_k} \frac{B_k\left(z^{-1}\right)}{A\left(z^{-1}\right)} u_k(i) + \frac{1}{A\left(z^{-1}\right)} e(i), \tag{4.12}$$

- *ARMAX*

$$y(i) = \sum_{k=1}^{q} z^{-d_k} \frac{B_k\left(z^{-1}\right)}{A\left(z^{-1}\right)} u_k(i) + \frac{C\left(z^{-1}\right)}{A\left(z^{-1}\right)} e(i), \tag{4.13}$$

- *FIR*

$$y(i) = \sum_{k=1}^{q} z^{-d_k} B_k\left(z^{-1}\right) u_k(i) + e(i), \tag{4.14}$$

- *OE*

$$y(i) = \sum_{k=1}^{q} z^{-d_k} \frac{B_k\left(z^{-1}\right)}{F_k\left(z^{-1}\right)} u_k(i) + e(i), \tag{4.15}$$

- *BJ*

$$y(i) = \sum_{k=1}^{q} z^{-d_k} \frac{B_k\left(z^{-1}\right)}{F_k\left(z^{-1}\right)} u_k(i) + \frac{C\left(z^{-1}\right)}{D\left(z^{-1}\right)} e(i). \tag{4.16}$$

Jeśli natomiast rozpatrywany jest proces bez wejść (modelowany w postaci ciągu czasowego), to jego liniowy model stochastyczny można opisać za pomocą typów:

- *AR*

$$y(i) = \frac{1}{A\left(z^{-1}\right)} e(i), \tag{4.17}$$

- *MA*

$$y(i) = C\left(z^{-1}\right) e(i), \tag{4.18}$$

- *ARMA*

$$y(i) = \frac{C\left(z^{-1}\right)}{A\left(z^{-1}\right)} e(i). \tag{4.19}$$

Wyróżniono także typ, który nazwano *modelem ogólnym*, gdyż obejmuje on wszystkie wymienione wyżej typy modeli liniowych. Dalsza dyskusja problemu wyznaczenia struktury prowadzona będzie więc dla modelu *ogólnego MISO*, jako że każdy inny typ modelu skalarnego można traktować jako szczególny przypadek modelu ogólnego. W modelu ogólnym transmitancję k-tego toru sterowania definiuje się jako:

$$G_k(z^{-1}) = z^{-d_k} \frac{B_k\left(z^{-1}\right)}{A\left(z^{-1}\right) F_k\left(z^{-1}\right)}, \tag{4.20a}$$

gdzie d_k jest czasem opóźnienia w torze, natomiast transmitancję toru zakłócenia $H(z^{-1})$ w postaci:

$$H(z^{-1}) = \frac{C\left(z^{-1}\right)}{A\left(z^{-1}\right) D\left(z^{-1}\right)}. \tag{4.20b}$$

Stopień danego wielomianu oznaczany będzie w dalszym ciągu przedrostkiem *d*, np. *dA* oznacza stopień wielomianu $A(z^{-1})$.

Strukturę modelu ogólnego MISO stanowi zbiór liczb całkowitych określających czasy opóźnienia i stopnie poszczególnych wielomianów $\mu = (d, dA, dB, dF, dC, dD)$, gdzie *d*, *dB* i *dF* są odpowiednimi wektorami:

$$d = [\, d_1 \;\; d_2 \;\; \ldots \;\; d_q \,]^T \tag{4.21a}$$

$$dB = [\, dB_1 \;\; dB_2 \;\; \ldots \;\; dB_q \,]^T \tag{4.21b}$$

$$dF = [\, dF_1 \;\; dF_2 \;\; \ldots \;\; dF_q \,]^T. \tag{4.21c}$$

W przypadku modelu SISO *d*, *dB* i *dF* są skalarami. Stopnie wielomianów, które nie występują w danym typie modelu, uważa się za równe zero. Strukturę modelu ciągu czasowego natomiast tworzą stopnie *dA* i *dC*.

Cechą wspólną wszystkich przedstawionych wyżej typów jest modelowanie za pomocą białego szumu *e(i)* tego wszystkiego, czego nie można przewidzieć, natomiast wiedza *a priori* o procesie przedstawiona jest za pomocą odpowiedniej struktury torów sterowania i toru zakłócenia. We wspomnianym już systemie *MULTI-EDIP* przewidziano także możliwość identyfikacji tylko modeli *transmitancji torów sterowania*, bez modelowania zakłócenia [101]. Jedyne założenie dotyczące zakłócenia jest tu takie, że jest ono *nieskorelowane* z każdym z wejść u_k. W takim przypadku zamiast kryterium aproksymacji, opartego na metodzie największej wiarygodności *ML*, wykorzystywana jest metoda zmiennych pomocniczych (instrumentalnych) *IV*, por. dodatek A.

4.2.2. Wybór typu modelu liniowego

W zasadzie typ modelu powinien być określony przez użytkownika na podstawie wiedzy *a priori* o procesie, często subiektywnej, oraz założonego celu identyfikacji. Jest to o tyle istotny problem, że wybór typu determinuje własności modelu, a zarazem metodę estymacji jego parametrów. Z drugiej strony jednak trudno jest oczekiwać, aby można było jednoznacznie określić właściwy typ modelu, jako że na ogół nie jest możliwe opisanie rzeczywistego obiektu S dokładnie za pomocą równań matematycznych. Stąd wybór typu stanowi zazwyczaj pewien kompromis między elastycznością modelu a jego prostotą oraz prostotą metody jego identyfikacji. Ponadto poszczególne typy modeli są między sobą równoważne w tym sensie, że poprzez proste przekształcenia i pewne uproszczenia można sprowadzić typ bardziej złożony do prostszego. Na przykład, dzieląc w modelu *ARMAX* obie strony równania (4.13) przez wielomian $C(z^{-1})$ otrzymuje się model *ARX*, w którym jednak nowe wielomiany $A(z^{-1})$ i $B(z^{-1})$

będą teoretycznie nieskończonego stopnia. Ponieważ zawsze można tak dobrać wielomian $C(z^{-1})$, aby był on stabilny, por. np. [79], więc kolejne współczynniki wielomianów powstałych z dzielenia przez $C(z^{-1})$ będą coraz mniejsze. Oznacza to, że wpływ odległych wartości wejść czy wyjść na wartość wyjścia w chwili bieżącej będzie pomijalnie mały, a więc stopnie wielomianów mogą być bezpiecznie ograniczone. Podobne rozumowanie można przeprowadzić dla innych typów modeli. W [83] pokazano, że dla dużej liczby próbek *N* modele *ARX* odpowiednio wysokiego stopnia umożliwiają aproksymację systemu liniowego dowolnego typu.

Oczywiście, im prostszy typ modelu, tym więcej parametrów będzie potrzebnych do odpowiedniego opisu własności dynamicznych obiektu. Z tego względu modeli typu *ARX*, a zwłaszcza *FIR*, nie można uznać za oszczędne. Jednak tylko dla tych dwóch typów (oraz modelu *AR* dla ciągu czasowego) błąd predykcji można przedstawić w postaci formy liniowej względem parametrów, a więc skorzystać z bezpośredniej metody najmniejszych kwadratów, por. dodatek A.1. Jest to podstawowa zaleta modeli *ARX*, powodująca, że często warto dopuścić do pewnej „rozrzutności" przy parametryzacji modelu, uzyskując w zamian korzyści przy estymacji jego parametrów. Chodzi tu nie tylko o nakład obliczeń, które są wymagane w metodach estymacji przeznaczonych dla typów bardziej złożonych – tu problem jest w pewien sposób rozwiązywany przez dostępność do coraz wydajniejszych komputerów. Problemem poważniejszym może być istnienie wielu minimów lokalnych, w sytuacji gdy wskaźnik jakości w postaci sumy kwadratów błędów predykcji jest optymalizowany dla błędów będących nieliniową funkcją parametrów. Dotyczy to szczególnie rekurencyjnej metody błędu predykcji *RPE*, która wymaga filtracji danych przez wielomiany modelu uzyskane w poprzednim kroku, co może w pewnych sytuacjach prowadzić do powstania poważnych problemów numerycznych i w efekcie uniemożliwić uzyskanie jakichkolwiek rozsądnych wyników. Przykładem takich problemów może być identyfikacja obiektu elektroakustycznego, por. 4.3.

Tak więc podstawowe rozwiązanie, jakie można w tym zakresie zaproponować użytkownikowi systemu komputerowego wspomagania identyfikacji, polega na przedstawieniu zestawu porad ułatwiających mu podjęcie decyzji co do wyboru typu modelu. W przypadku identyfikacji modelu obiektu przykładowy zestaw takich informacji, wynikający między innymi z własnych doświadczeń autora, mógłby wyglądać następująco:

- Jeżeli nie ma podstaw do przyjęcia innego typu modelu, to identyfikację powinno się zaczynać od modeli najprostszych, tj. *ARX*, szczególnie jeżeli poziom zakłócenia *NSR* jest stosunkowo mały.
- Modele *ARX* i *ARMAX* najlepiej nadają się dla celów predykcji.

- Modele *FIR* są także proste do identyfikacji i zawsze są stabilne, ale do prawidłowego modelowania procesu wymagają znacznie większej liczby parametrów niż modele *ARX*, stąd nie są oszczędne. Liczbę współczynników potrzebnych do modelowania dynamiki procesu można oszacować korzystając z modelu odpowiedzi impulsowej otrzymanego metodą nieparametryczną, por. [101].
- Parametry modeli *ARX*, *ARMAX* oraz *FIR* zawsze dadzą się wyznaczyć (o ile tylko zadanie jest wystarczająco dobrze uwarunkowane numerycznie). Dla pozostałych typów modeli metoda estymacji może nie dać wyników, szczególnie wtedy, gdy w modelu występują bieguny bliskie okręgu jednostkowego.
- Modele *OE* dają zazwyczaj najlepsze wyniki jeśli za kryterium jakości modelu przyjąć dokładność modelowania toru sterowania. Jest to efekt minimalizacji sumy kwadratów błędów wyjścia.
- Jeżeli poziom zakłócenia *NSR* jest stosunkowo duży oraz wiadomo, że własności dynamiczne tego zakłócenia różnią się od własności dynamicznych występujących w torze sterowania, to najbardziej odpowiedni jest model *BJ*.
- Jeżeli poziom zakłócenia jest duży i nie ma potrzeby modelowania tego zakłócenia, to najlepiej wybrać model w postaci *transmitancji toru sterowania*.
- Struktura *modelu ogólnego* jest zbyt złożona dla większości praktycznych przypadków. Wprowadzona została raczej jako teoretyczne uogólnienie wszystkich możliwych struktur modeli liniowych oraz w celu testowania.

W sytuacji identyfikacji modelu ciągu czasowego można podać podobny zestaw reguł, z tym że tutaj można także posiłkować się analizą wykresów funkcji autokorelacji oraz korelacji cząstkowej, por. [19].

4.3. Struktura modelu dla obiektu elektroakustycznego

W rozpatrywanym przypadku identyfikacji obiektu elektroakustycznego modele potrzebne są dla celów filtracji, por. 2.3.1, więc najwygodniej jest tu zastosować opis wejściowo-wyjściowy (model w postaci „czarnej skrzynki"). Ponieważ filtracja w układzie ATH musi być wykonywana w czasie rzeczywistym, więc modele te powinny być odpowiednio proste, najlepiej *liniowe*. Z drugiej jednak strony ich struktura musi być odpowiednio bogata, aby umożliwić wystarczająco dokładną aproksymację skomplikowanej dynamiki identyfikowanych obiektów.

Ponieważ rozpatrywany obiekt *MIMO* można traktować jako 4 niezależne (aczkolwiek podobne) obiekty *MISO* o 3 wejściach (sygnałach sterujących głośnikami, generowanych przez kartę *DSP*) oraz 1 wyjściu (sygnał z odpowiedniego mikrofonu dostarczany do karty *DSP*), por. 2.3.2, więc w dalszym ciągu będzie rozważany problem identyfikacji pojedynczego obiektu *MISO*, którego model liniowy opisuje równanie (4.9) dla $q = 3$.

Ponieważ każdy z torów składa się z podobnych elementów, jak przetworniki C/A, filtry, wzmacniacze itp., wydaje się uzasadnione modelowanie obiektu za pomocą struktury *ARX*:

$$A\left(z^{-1}\right)y(i)=\sum_{k=1}^{3} z^{-dk} B_k\left(z^{-1}\right)u_k(i)+e(i), \tag{4.22}$$

gdzie

$$A\left(z^{-1}\right)=1+a_1 z^{-1}+a_2 z^{-2}+\ldots+a_{dA} z^{-dA}, \tag{4.23}$$

$$B_k\left(z^{-1}\right)=b_{k,0}+b_{k,1} z^{-1}+b_{k,2} z^{-2}+\ldots+b_{k,dB_k} z^{-dB_k}. \tag{4.24}$$

Struktura modelu określona jest więc przez zbiór liczb całkowitych:

$$\mu=\left(\left[d_1\ d_2\ d_3\right],\ dA,\ \left[dB_1\ dB_2\ dB_3\right]\right), \tag{4.25}$$

a wektor parametrów θ zawiera współczynniki poszczególnych wielomianów:

$$\theta=[\,a_1\ a_2\ \ldots\ a_{dA}\quad b_{1,0}\ b_{1,1}\ \ldots\ b_{1,dB_1}\quad b_{2,0}\ b_{2,1}\ \ldots\ b_{2,dB_2}\quad b_{3,0}\ b_{3,1}\ \ldots\ b_{3,dB_3}\,]^{\mathrm{T}} \tag{4.26}$$

Zadanie wyznaczenia zbioru wskaźników struktury (4.25) oraz estymacji zbioru parametrów (4.26) omówiono w punkcie 5.3.

Transmitancja k-tego toru sterowania jest równa

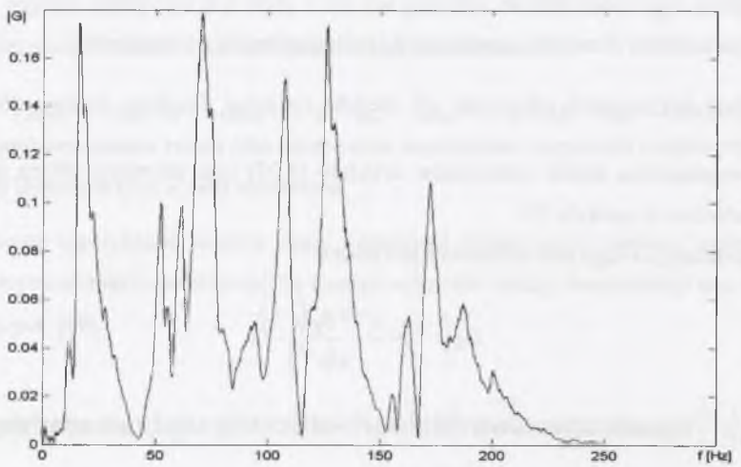
$$G_k\left(z^{-1}\right)=\frac{z^{-d_k} B_k\left(z^{-1}\right)}{A\left(z^{-1}\right)}, \tag{4.27}$$

gdzie $G_k\left(z^{-1}\right)$ oznacza odpowiednio S_{kl} (dla $l = 1, 2, 3$) lub F_k. Z pewnym przybliżeniem można powiedzieć, że wielomian $B_k\left(z^{-1}\right)$ przedstawia własności specyficzne dla toru sterowania od k-tego głośnika, a $A\left(z^{-1}\right)$ wspólne dla wszystkich torów.

Za wyborem typu *ARX* przemawia jeszcze inny argument. W przypadku modelu *ARX* błąd predykcji $\varepsilon(i \mid i-1)$ można przedstawić w postaci liniowej funkcji parametrów:

$$\varepsilon(i \mid i-1)=A\left(z^{-1}\right)\left[y(i)-\sum_{k=1}^{3} B_k\left(z^{-1}\right)u_k\left(i-d_k\right)\right], \tag{4.28}$$

co umożliwia estymację tych parametrów przy wykorzystaniu klasycznej metody najmniejszych kwadratów. Natomiast w przypadku wyboru innych struktur modeli, jak np. *ARMAX*, *OE*, *BJ* itd. minimalizacja funkcji strat wymaga stosowania metod optymalizacji nieliniowej, np. w algorytmie *RPE* jest to metoda Newtona-Raphsona, por. dodatek A. Metody te wymagają jednak filtracji danych przy użyciu ocen parametrów uzyskanych w trakcie działania algorytmu. Jeśli spojrzeć na wykres typowej zależności amplitudy od częstotliwości w dowolnej ścieżce elektroakustycznej, to można zauważyć wiele dolin i pików, por. rys. 4.1. Świadczy to o tym, że w modelu transmitancyjnym powinno występować wiele zer (odpowiadających dolinom) i biegunów (odpowiadających pikom) bliskich okręgu jednostkowego. To, niestety, prowadzi do niestabilności numerycznej wszystkich metod opartych na *RPE*. Podobnie dzieje się także w przypadku metody zmiennej pomocniczej *IV*. Testowano tu zarówno system *MULTI-EDIP*, jak i *System Identification Toolbox* i w żadnym przypadku nie udało się uzyskać modelu, poza strukturami *ARX* lub *FIR*. Ponadto metody te w sytuacji tak wysokich stopni wielomianów, jakie są tu wymagane, byłyby zbyt czasochłonne. Poza tym, jak pokazano w [83], dla dużej liczby próbek N modele *ARX* odpowiednio wysokiego stopnia umożliwiają aproksymację systemu liniowego dowolnego typu.



Rys. 4.1. Przykładowa charakterystyka amplitudowa obiektu
Fig. 4.1. Exemplary magnitude of frequency response of the plant

Gdyby modelować obiekt jako zbiór modeli *SISO*, niezależnie dla każdego głośnika, to w równaniu (4.26) wystąpiłby dla każdego toru inny wielomian $A_k\left(z^{-1}\right)$, a struktura modelu *SISO* dla k-tego toru sterowania byłaby określona jako trójka $\left(d_k, dA_k, dB_k\right)$.

Do opisu dynamiki można stosować także modele prostsze, tzn. typu *FIR*, zawierające tylko wielomiany $B_k\left(z^{-1}\right)$, ale przy tak złożonej dynamice obiektu, aby uzyskać odpowiednio dokładny opis musiałyby być to wielomiany o stopniu większym niż 200, por. [90]. O ile więc można rozważać stosowanie modeli *FIR* dla obiektów *SISO*, to dla obiektu *MISO* podejście takie wymagałoby rozwiązania układu równań o zbyt dużej liczbie niewiadomych.

Rozdział 5

Procedura identyfikacji

Zgodnie z uwagami z rozdziału 2 istotą identyfikacji jest wnioskowanie na podstawie danych wynikające z metodologii eksperymentalnej metody badawczej i ogólnej procedury poprawiania i falsyfikacji hipotez. W metodzie tej definiuje się „prawdę" jako taką *hipotezę*, która nigdy nie zostanie sfalsyfikowana przez jakąkolwiek obserwację. Ponieważ w praktyce nie jest to możliwe, stąd procedura ta w ogólności nigdy nie powinna się zakończyć. Jednak gdy model tworzony jest dla celów inżynierskich, procedura identyfikacji może zostać zatrzymana po uzyskaniu modelu, który można uznać za *odpowiedni* do tego celu, por. [17].

Po omówieniu zasad poprawnego wykonania eksperymentu oraz modelowania polegającego na wyborze struktury modelu $\mathcal{M}$ tematem kolejnego rozdziału jest dyskusja szczegółów dotyczących wykonywania procedury identyfikacji. Aby powiązać to, co należy wyznaczyć (tj. model), z tym, co obserwuje się w doświadczeniu (dane), można wykorzystać albo podejście bayesowskie, albo wynikające z niego wnioskowanie statystyczne, którego celem jest wyznaczenie *wiarygodności* tej „prawdy". W punkcie 5.1 przedstawiono podstawowe zadania występujące w procedurze, tj. zadania aproksymacji, falsyfikacji i walidacji modelu, po czym w punkcie 5.2 przedstawiono tę procedurę w postaci sformalizowanej. Omówiono także możliwe przyczyny powstawania błędów oraz ewentualne inne źródła informacji i możliwości komputerowego wspomagania w zakresie określania struktury modelu. W następnych punktach przedstawiono szczegóły dotyczące komputerowego wspomagania w realizacji procedury dla modeli liniowych, takie jak kryterium wyboru, opis algorytmu poszukiwania struktury czy testowanie i analiza otrzymanego modelu.

5.1. Zadania aproksymacji, falsyfikacji i walidacji

Identyfikacja powinna się zakończyć wybraniem jakiegoś konkretnego modelu $\hat{M}$ należącego do założonego zbioru $\mathcal{M}$. Jednocześnie model ten powinien spełniać dwa podstawowe wymagania: powinien być *odpowiedni do celu* oraz *niesprzeczny z danymi*. W klasycznym ujęciu realizowane jest to w ten sposób, że najpierw określa się pewne *kryterium aproksymacji*, które umożliwia wyznaczenie $\hat{M}$, a następnie dokonuje się jego *walidacji*, por. np. [79, 115, 132]. Tu warto jednak zauważyć, że to, co zwykle się uważa za walidację, jest

w zasadzie sprawdzeniem, czy model odpowiednio dobrze opisuje dane, a więc zgodnie z definicjami z punktu 2.1.2 jest to właściwie zadanie *falsyfikacji* modelu. Na ogół zakłada się także milcząco, że przyjęte kryterium aproksymacji jest najbardziej odpowiednie do celu.

5.1.1. Kryterium aproksymacji a cel identyfikacji

Wybór konkretnego modelu ze zbioru $\mathcal{M}$ wymaga określenia reguły aproksymacji $\mathcal{A}: (\mathbf{U}_N, \mathbf{Y}_N) \to \hat{M}$, która zwykle jest formułowana w postaci kryterium jakości modelu. W przypadku omawianej już metody największej wiarygodności *ML* reguła aproksymacji polega na minimalizacji funkcji strat V wynikającej z równań (2.10) i (4.2):

$$\hat{M} = \arg\min\{-\ln L(M \mid \mathbf{U}_N, \mathbf{Y}_N) / N \mid M \in \mathcal{M}\}. \tag{5.1}$$

Nie wszystkie metody estymacji parametrów opierają się na kryterium, które jest jawnie określone, np. metoda zmiennych pomocniczych (instrumentalnych) *IV* (*Instrumental Variable*), por. np. [131]. U podstaw tego typu metod leżą głównie łatwość obliczeń lub statystyczne własności zbieżności ocen wyznaczanych dla długich ciągów danych.

Kryterium wyboru (tj. reguła aproksymacji $\mathcal{A}$) stanowi element występujący w zadaniu identyfikacji, który powinien być określony *a priori*, podobnie jak struktura $\mathcal{M}$. Jednak jeśli cel identyfikacji Q jest dobrze zdefiniowany, to naturalnym kryterium wynikającym w sposób jednoznaczny z tego celu powinien być wybór takiego modelu, który jest *najbardziej odpowiedni* do tego celu. Szczególny przypadek takiego podejścia stanowi powszechnie stosowane kryterium aproksymacji oparte na *minimalizacji sumy kwadratów błędów predykcji*, stanowiące istotę tzw. metod błędu predykcji (*Prediction-Error Methods*, *PEM*), por. np. [79, 132]. Kryterium takie jest odpowiednie dla celów prognozowania i sterowania, albo gdy celem tym jest uzyskanie modelu, którego odpowiedź dobrze symuluje odpowiedź obiektu. W takim przypadku można przyjąć, że kryterium $\mathcal{A}$ definiuje także cel identyfikacji Q, z tym że wtedy samo to kryterium nie może stanowić jedynej podstawy do potwierdzenia wystarczającej dokładności modelu – potrzebne są jeszcze dodatkowe testy walidacji wybranego modelu, por. 5.1.3.

Jeżeli jednak przyjęto taki cel identyfikacji Q, który niekoniecznie musi być związany z kryterium stosowanym w *PEM*, to wtedy należałoby wybrać takie kryterium $\mathcal{A}$, które byłoby bardziej odpowiednie dla tego celu Q. Jeżeli np. celem identyfikacji jest uzyskanie modelu o odpowiednio dokładnej charakterystyce częstotliwościowej, to zamiast minimalizacji błędu predykcji jako kryterium aproksymacji można zastosować minimalizację sumy kwadratów błędów zdefiniowanych w dziedzinie częstotliwości, por. [139]. Jednak ponieważ nie zawsze

jest to łatwe, stąd często stosowanym rozwiązaniem jest korzystanie z *PEM*, które są dobrze opracowane i stosunkowo proste w implementacji (zwłaszcza dla modeli liniowych), a następnie przeprowadzenie walidacji modelu dla testu zdefiniowanego w taki sposób, aby był odpowiedni dla wyznaczonego celu Q. Przykładem takiego podejścia jest identyfikacja modelu obiektu elektroakustycznego, por. 5.1.4.

Podsumowując, przyjęcie modelu dynamicznego w postaci modelu stochastycznego o zadanej strukturze $\mathcal{M} = (\mathcal{F}, \mu)$ oraz kryterium aproksymacji $\mathcal{A}$ opartego na funkcji wiarygodności umożliwia estymację parametrów θ modelu na drodze minimalizacji funkcji strat:

$$\hat{\theta} = \arg\min V(\theta \mid \mathcal{F}, \mu, \mathbf{U}_N, \mathbf{Y}_N), \tag{5.2}$$

gdzie $V(\theta \mid \mathcal{F}, \mu, \mathbf{U}_N, \mathbf{Y}_N)$ zdefiniowane jest przez (4.2). Dla liniowych modeli wejściowo-wyjściowych zadanie aproksymacji (5.2) można sprowadzić do minimalizacji sumy kwadratów błędów predykcji, por. 4.2, a więc można przyjąć, że jest ono odpowiednie do celów związanych ze sterowaniem, predykcją czy symulacją.

5.1.2. Falsyfikacja modelu

Zasadniczo *falsyfikacja* w zadaniu identyfikacji oznacza możliwość sprawdzenia, czy model M nie jest *sprzeczny* z danymi $(\mathbf{U}_N, \mathbf{Y}_N)$. Zadanie to jest stosunkowo proste w sytuacji, gdy dane pochodzą z systemu deterministycznego, gdyż wtedy wykorzystując np. symulację modelu można sprawdzić, czy obliczone wyjście y zgadza się z danym $\mathbf{y}$. W przypadku systemu losowego samo porównanie y z $\mathbf{y}$ nie wystarczy do stwierdzenia, o ile M różni się od $\mathbf{S}$, bowiem nawet dla $M = \mathbf{S}$ wyjścia nie będą odtwarzane dokładnie, a błąd tego odtworzenia będzie zależny od procesu losowego e. W związku z tym, oprócz porównywania wyjść w celu określenia, co należy uważać za model *niesfalsyfikowany*, powinno się także korzystać z kryteriów, które wykorzystują teorię statystycznego testowania hipotez, por. np. [76].

W metodach tych zakłada się często statystyczną niezależność między modelem a danymi użytymi do testowania. Wymaga to podzielenia ciągu danych na dwa zbiory: zbiór *danych do estymacji* oraz zbiór *danych do testowania*, albo pozyskania nowych danych. Podejście takie bywa nazywane falsyfikacją skrośną (*cross-falsification*, por. [17], albo też *cross-validation*, jeśli nie rozróżnia się wyraźnie testów falsyfikacji i walidacji, jak np. w [79]). Wadą tego rozwiązania jest to, że nie można wykorzystać całej informacji do wyznaczenia modelu, gdyż część danych musi być przeznaczona do testowania.

Falsyfikację modelu można także przeprowadzić dla tych samych danych, których użyto do wyznaczenia modelu, korzystając z odpowiednich testów: bezwarunkowych lub warunkowych.

Testy *bezwarunkowe* polegają na sformułowaniu hipotezy zerowej $\mathcal{H}_0$ typu „model M jest *równoważny* obiektowi $\mathbf{S}$" oraz określeniu dziedziny $\Sigma(\gamma \mid M)$ w przestrzeni danych takiej, że prawdopodobieństwo tego, że dane wygenerowane przez M będą należały do $\Sigma(\gamma \mid M)$, ma zadaną wysoką wartość poziomu ufności γ. Jeżeli okaże się, że $(\mathbf{U}_N, \mathbf{Y}_N) \notin \Sigma(\gamma \mid M)$, to odrzuca się hipotezę $\mathcal{H}_0$, a więc uznaje model M za *sfalsyfikowany*. Jednak zamiast definiować dziedzinę $\Sigma(\gamma \mid M)$ bezpośrednio w przestrzeni danych, łatwiej jest to zrobić w sposób pośredni, korzystając z tzw. statystyk bezparametrowych [17].

W przypadku modeli stochastycznych hipotezą podlegającą falsyfikacji może być ciąg innowacji (4.6), który powinien być gaussowski i ortogonalny. Ponieważ w rzeczywistości ciąg ten jest niedostępny, więc testuje się funkcje kowariancji obliczonych reszt (błędów predykcji) [133]. W literaturze proponowanych jest wiele różnych testów, które noszą ogólną nazwę testów białości reszt (*whiteness of residuals*), por. np. [79, 86]. Ponieważ błędy te powinny stanowić ciąg zmiennych nieskorelowanych i jednocześnie być nieskorelowane z poprzednimi wartościami wejść, więc najwygodniejszym sposobem testowania jest badanie funkcji *autokorelacji reszt* oraz *korelacji wzajemnej reszt i wejść*. Ponadto testując skorelowanie reszt z przyszłymi wartościami wejść można testować hipotezę, że eksperyment został przeprowadzony w układzie otwartym. Ze względu na to, że do wyznaczenia reszt wykorzystuje się obliczony model $\hat{M}$, a nie rzeczywisty obiekt $\mathbf{S}$, więc poziom ufności γ powinien być przyjęty jako odpowiednio wysoki (albo inaczej poziom ryzyka odrzucenia modelu poprawnego $\alpha = 1 - \gamma$ odpowiednio mały), gdyż w przeciwnym razie testy te mogą mieć tendencję do nadparametryzacji, por. [133]. Stąd typowe podejście, por. np. [101], polega na przyjęciu granicy 3 odchyleń standardowych oszacowanego błędu wyznaczenia funkcji autokorelacji białego szumu zgodnie z uproszczonym wzorem Bartletta [9]. Testy autokorelacji reszt oraz korelacji wzajemnej reszt z ciągiem pobudzeń nie umożliwiają jednak wykrycia nieliniowości w identyfikowanym procesie, stąd jeśli brak jest pewności co do poprawności modelowania liniowego, to można stosować także testy korelacji badające trzecie momenty [77].

Testy *warunkowe*, nazywane w literaturze także testami istotności, wymagają z kolei postawienia hipotezy alternatywnej $\mathcal{H}_1$. Testy tego typu szczególnie nadają się do warunkowej falsyfikacji struktur na drodze badania statystyk porównujących modele o różnych strukturach. Ponieważ hipoteza $\mathcal{H}_1$ stanowi założenie i nie jest testowana, więc jeśli będzie ona fałszywa, to będzie stanowić „pułapkę". Na ogół przyjęcie struktury alternatywnej jako zbyt złożonej nie stanowi potencjalnego źródła błędów, natomiast niedoszacowanie jej złożoności może prowadzić do powstania pułapki w sytuacji, gdy brak jest możliwości walidacji modelu, por. [17]. Dla bardzo długich ciągów próbek może być stosowany test stosunku wiarygodności (*Likelihood-Ratio*, *LR*), por. np. [142], jednak w praktyce, dla ograniczonej liczby danych N

zamiast testu *LR* stosuje się raczej związany z nim *test F*, w którym porównuje się wartości funkcji strat obliczone dla dwóch struktur modeli, por. [130]. *Test F*, który w przypadku modeli dynamicznych jest właściwie asymptotycznym testem χ^2, stosowany jest głównie do wyboru struktury i jest ściśle związany z innymi kryteriami wyboru, por. 5.3.1.

Jako uzupełnienie różnych testów statystycznych warto jest także w zadaniu falsyfikacji wykorzystać porównanie *różnych charakterystyk* modeli wyznaczonych *różnymi metodami* dla *różnych fragmentów* ciągów danych, por. [56].

Wszystkie tego typu kryteria można wyrazić za pomocą pewnej miary *odległości* $\Gamma(M \mid \mathbf{U}_N, \mathbf{Y}_N)$ między modelem a danymi, oraz przyjętej *granicznej* wartości γ dla akceptacji modelu. W wyniku takiego podejścia można zdefiniować parametryczną rodzinę zbiorów modeli *niesfalsyfikowanych* $\mathcal{N}$ jako:

$$\mathcal{N}(\gamma \mid \mathbf{U}_N, \mathbf{Y}_N) = \{M \mid \Gamma(M \mid \mathbf{U}_N, \mathbf{Y}_N) \le \gamma\}. \tag{5.3}$$

Odrzucenie modelu oznacza więc, że $M \notin \mathcal{N}(\gamma \mid \mathbf{U}_N, \mathbf{Y}_N)$ lub równoważnie $\Gamma(M \mid \mathbf{U}_N, \mathbf{Y}_N) > \gamma$, gdzie γ jest dane.

Warto tu zauważyć, że skoro aproksymacja oparta na zasadzie największej wiarygodności zapewnia najlepsze dopasowanie modelu do danych, to jeżeli model *ML* zostanie sfalsyfikowany jako niezgodny z danymi, to jednocześnie falsyfikowana jest cała struktura $\mathcal{M}$.

Należy także pamiętać, że zbiór modeli $\mathcal{N}$ opisuje dane $(\mathbf{U}_N, \mathbf{Y}_N)$ pochodzące z konkretnego doświadczenia, ale niekoniecznie opisuje system, który te dane generuje. A więc pewność, że zachowanie systemu będzie takie samo dla powtórzonego doświadczenia lub że obiekt poddany innemu pobudzeniu będzie opisany przez ten sam model, musi wynikać z innych przesłanek, a nie tylko samych danych, por. punkt 3.1 na temat wymagań wobec eksperymentu identyfikacyjnego.

5.1.3. Walidacja modelu

Jeżeli cel identyfikacji Q jest wyraźnie określony i da się zdefiniować zbiór modeli *odpowiednich do celu*, to do zatrzymania iteracyjnej procedury identyfikacji można wykorzystać testy *walidacji* modelu. Na ogół zbiór modeli odpowiednich do celu Q zależy od obiektu **S**, np. wtedy, gdy celem tym jest sterowanie obiektem, ale niekoniecznie, np. gdy celem identyfikacji jest kodowanie danych w celu zmniejszenia kosztów transmisji do odbiorcy [136]. Jednocześnie zbiór Q zależy od tego, w jaki sposób model będzie wykorzystany, a z tym wiążą się dwa podstawowe wymagania, tj. by model był *reprezentatywny* i *dokładny*, por. 2.1.1. W przybliżeniu oznacza to, że model powinien w odpowiedni sposób reagować na pobudzenie

występujące na etapie jego stosowania (a pobudzenie to może mieć inne własności niż to, które wystąpiło w czasie eksperymentu identyfikacyjnego), przy czym odpowiedź ta powinna być *wystarczająco dokładna*. Wymagana *dokładność* odpowiedzi zależy od celu – procedura identyfikacji może zostać zatrzymana wtedy, gdy dokładność jest już wystarczająca z punktu widzenia zastosowania modelu. Warunek *reprezentatywności* jest bardziej złożony i jego spełnienie jest zapewnione przede wszystkim przez spełnienie warunków dostatecznego pobudzenia i powtarzalności, por. punkt 3.1.

Zadania walidacji modelu nie można rozwiązać bezpośrednio za pomocą zbioru modeli *odpowiednich do celu* $Q(\mathbf{S})$, skoro cel ten zależy od nieznanego obiektu. Bohlin w [17] proponuje tu skorzystanie ze wzoru Bayesa oraz statystycznej zależności pomiędzy nieznanym źródłem (**R**, **S**) a znanymi danymi $(\mathbf{U}_N, \mathbf{Y}_N)$ i oparcie reguły walidacji na prawdopodobieństwie jako parametrze. Opierając kryterium walidacji na prawdopodobieństwie

$$\Sigma(M \mid \mathbf{U}_N, \mathbf{Y}_N) \equiv \Pr\{M \in Q(\mathbf{S}) \mid \mathbf{U}_N, \mathbf{Y}_N\}, \tag{5.4}$$

oraz przyjmując, że prawdopodobieństwo to powinno być większe niż założona wartość graniczna σ, można zdefiniować parametryczną rodzinę zbiorów modeli *odpowiednich do celu* jako:

$$Q(\sigma \mid \mathbf{U}_N, \mathbf{Y}_N) = \{M \mid \Sigma(M \mid \mathbf{U}_N, \mathbf{Y}_N) \ge \sigma\}, \tag{5.5}$$

analogicznie jak to zrobiono dla zbioru modeli *niesfalsyfikowanych*.

Jednak w praktyce wyznaczenie (5.4) jest bardzo trudne do wykonania i stąd zadanie walidacji definiuje się raczej w sposób heurystyczny, korzystając ze znajomości konkretnego zadania identyfikacji. Jeśli można zdefiniować model **M** *równoważny obiektowi ze względu na cel*, tj. taki, że $Q(\mathbf{S}) = Q(\mathbf{M})$, to wtedy zamiast posługiwać się prawdopodobieństwem, kryterium walidacji można sformułować za pomocą pewnej miary odległości między modelem M a modelem *równoważnym obiektowi ze względu na cel*: $\Delta(M \mid \mathbf{M})$. Tak więc zbiór Q modeli *odpowiednich do celu* można zdefiniować jako

$$Q(\delta \mid \mathbf{M}, \mathbf{U}_N, \mathbf{Y}_N) = \{M \mid \Delta(M \mid \mathbf{M}, \mathbf{U}_N, \mathbf{Y}_N) \le \delta\}, \tag{5.6}$$

gdzie δ jest parametrem określającym granicę dokładności modelu.

W przypadku badań symulacyjnych sprawa jest prosta, gdyż znając obiekt **S** można przyjąć po prostu **M** = **S**. W przypadku identyfikacji modelu obiektu rzeczywistego można próbować zdefiniować **M** jako model, który uzyskano inną metodą niż podlegający walidacji model M. Podejście to wykorzystano do walidacji modelu obiektu elektroakustycznego. Inne stosowane często rozwiązanie polega na minimalizacji tzw. *kryterium informacyjnego*, bazującego na odległości informacyjnej między prawdziwą a estymowaną gęstością prawdopodobieństwa, co

omówione zostało bliżej w punkcie 5.3.1. Zgodnie z zasadą prostoty celem tego kryterium jest uzyskanie pewnego kompromisu między złożonością modelu a dążeniem do minimalizacji *wariancji błędów predykcji.*

5.1.4. Walidacja w identyfikacji obiektu elektroakustycznego

Podstawowym celem identyfikacji jest tu otrzymanie modeli torów elektroakustycznych w postaci transmitancyjnej, umożliwiającej odpowiednią filtrację sygnałów w układzie ATH, por. 2.3.1. Ponieważ rozpatrywany układ oparty jest na zasadzie kompensacji, więc dokładność otrzymanych modeli ma bardzo duży wpływ na wyniki tłumienia. W szczególności dotyczy to dokładności charakterystyk częstotliwościowych, jako że błędy wyznaczenia amplitudy obniżają szybkość zbieżności algorytmu adaptacji, a błędy fazy mogą nawet powodować powstanie niestabilności w układzie [47]. Błędy te nie powinny przekraczać wartości $\pm\pi/2$, por. [26, 72]. Ponadto w przypadku błędów modelowania sprzężenia akustycznego możliwa jest destabilizacja układu lub powstanie zjawiska chaosu, por. [31]. Stąd poszukiwany jest model parametryczny M w postaci transmitancji potrzebnych do zaprojektowania odpowiednich filtrów, który jednocześnie będzie *odpowiednio dokładny w dziedzinie częstotliwości*, aby zapewnić poprawną pracę układu ATH i wymaganą jakość tłumienia. Pojawia się więc istotny problem zdefiniowania zbioru Q modeli odpowiednich do tego celu, czyli określenie kryterium umożliwiającego sprawdzenie, czy uzyskany model jest odpowiednio dokładny, zanim jeszcze zostanie poddany ostatecznej akceptacji w wyniku jego zastosowania.

Aby móc dokonać walidacji modelu, należy przede wszystkim określić, co należy rozumieć przez model $\mathbf{M}$ *równoważny do obiektu ze względu na cel*, a następnie zdefiniować pewną miarę odległości od niego modelu $\hat{M}$, który wyznaczono metodą *PEM*. Tutaj, z braku innych możliwości, wydaje się być uzasadnione przyjęcie za $\mathbf{M}$ charakterystyki amplitudowo-fazowej wyznaczonej na drodze klasycznej analizy widmowej i porównywanie z nią charakterystyki obliczonej dla otrzymanego modelu parametrycznego.

Jako kryterium walidacji zaproponowano więc ocenę dokładności modelu opartą na porównaniu dla każdego toru sterowania charakterystyki częstotliwościowej $\hat{G}_k(j\omega)$ obliczonej dla modelu $\hat{G}_k(z^{-1})$ ze wzorcem $\mathbf{G}_k(j\omega)$, który otrzymano metodą nieparametryczną. Aby zwiększyć pewność, że wzorce $\mathbf{G}_k(j\omega)$ dobrze aproksymują własności obiektu, na potrzeby tej pracy przyjęto, że wyznaczone powinny być one na podstawie dodatkowych doświadczeń, w których kolejno każdy z głośników pobudzany będzie sygnałem wielosinusoidalnym o pełnym widmie. Charakterystyki poszczególnych ścieżek można obliczyć korzystając

z empirycznego estymatora *ETFE* [79]. Testowanie modelu dla innego zbioru danych niż ten, który był użyty do estymacji, nosi nazwę walidacji skrośnej (ang. *cross-validation*) i jest zalecane w literaturze, por. 5.1.2.

Tak więc model parametryczny $\hat{G}_k(z^{-1})$ uznaje się za *odpowiedni do celu*, jeśli jego charakterystyka amplitudowo-fazowa jest *wystarczająco dokładna* w porównaniu ze wzorcem. Należy jeszcze sprecyzować, co to znaczy „wystarczająco dokładna". Można to np. wyrazić za pomocą pewnej miary odległości Δ pomiędzy modelem $\hat{G}_k(j\omega)$ a wzorcem $\mathbf{G}_k(j\omega)$ obliczonej dla częstotliwości względnych ω, będących wielokrotnościami częstotliwości podstawowej $\Omega = 2\pi/N$ i przyjętej *a priori* granicy dokładności δ. Wtedy przez model odpowiedni do celu rozumie się taki, który spełnia warunek $\Delta[\hat{G}_k(j\omega), \mathbf{G}_k(j\omega)] \le \delta$, por. 5.1.3.

Niestety, na obecnym etapie wiedzy bardzo trudno jest określić odpowiednią wartość δ, która zapewniałaby właściwe działanie układu ATH. Stąd w celu walidacji modelu proponuje się obliczać błąd fazy:

$$\Delta\varphi_k(\omega) = \left| \operatorname{Arg}(\hat{G}_k(j\omega)) - \operatorname{Arg}(\mathbf{G}_k(j\omega)) \right| \tag{5.7}$$

i sprawdzać, czy nie przekracza on granicy $\pi/2$ dla częstotliwości ω w przewidywanym zakresie pracy, zgodnie z przedstawionymi wyżej uwagami.

Jednocześnie definiuje się wskaźnik dokładności charakterystyki amplitudowej w postaci błędu średniokwadratowego:

$$D_k = \sqrt{\frac{1}{l_2 - l_1 + 1} \sum_{l=l_1}^{l_2} \left(\left| \hat{G}_k(j\Omega l) \right| - \left| \mathbf{G}_k(j\Omega l) \right| \right)^2}, \tag{5.8}$$

gdzie Ωl_1 i Ωl_2 odpowiadają dolnej i górnej granicy częstotliwości pracy. Wskaźnik ten wykorzystuje się do porównania modeli uzyskanych dla różnych struktur.

Tu oczywiście można by postawić pytanie, czy zamiast przyjmować kryterium aproksymacji $\mathcal{A}$ oparte na minimalizacji błędu w dziedzinie czasu nie należałoby raczej zmienić tego kryterium na bardziej dopasowane do celu identyfikacji. Strukturę modelu μ oraz jego parametry θ można by dobierać tak, aby minimalizować funkcję strat definiowaną w dziedzinie częstotliwości, np. w postaci normy $\ell[\hat{G}_k(j\omega), \mathbf{G}_k(j\omega)]$ zdefiniowanej dla różnicy między charakterystyką amplitudowo-fazową modelu parametrycznego a modelu uzyskanego metodą nieparametryczną, tak jak to zaproponowano w [139]. O ile jednak tego typu podejście jest uzasadnione w przypadku identyfikacji typowych obiektów będących filtrami dolnoprzepustowymi niskiego rzędu, to w tym przypadku, gdy mamy do czynienia z obiektami

elektroakustycznymi o tak złożonej dynamice, metoda ta nie byłaby odpowiednia ze względów numerycznych.

5.2. Iteracyjny charakter procedury

W identyfikacji można wyróżnić następujące zadania:

- zaplanowanie eksperymentu identyfikacyjnego (**R, S, T**) i jego wykonanie, tj. zebranie danych ($\mathbf{U}_N$, $\mathbf{Y}_N$), co omówiono w rozdziale 3,
- modelowanie: ustalenie struktury modelu $\mathcal{M} = (\mathcal{F}, \mu)$, por. rozdział 4,
- aproksymacja: wyznaczenie modelu $\hat{M}$ dla założonej struktury, por. 5.1.1,
- testowanie: walidacja i falsyfikacja otrzymanego modelu, por. 5.1.2 i 5.1.3,

które składają się na procedurę identyfikacji. Akceptacja modelu kończy etap identyfikacji i można przejść do etapu stosowania modelu.

Niestety, rzadko kiedy udaje się uzyskać odpowiedni model po jednym przejściu przedstawionej sekwencji kroków. Na ogół brak jest odpowiedniej wiedzy *a priori*, aby można było od razu zaplanować prawidłowy eksperyment **R** i przyjąć właściwą strukturę $\mathcal{M}$, stąd procedura identyfikacji jest zazwyczaj procesem iteracyjnym, w którym występuje konieczność powtarzania kolejnych etapów. Sprawę utrudnia fakt, że niektóre z etapów identyfikacji są ze sobą ściśle powiązane tworząc dodatkowe sprzężenia. Dotyczy to przede wszystkim planowania eksperymentu i modelowania. Z jednej strony planując eksperyment powinno się już znać najogólniejszą strukturę modelu $\mathcal{M}_{max}$, jaka będzie badana, gdyż od tego mogą zależeć warunki dostatecznego pobudzenia obiektu, np. zmienność sygnału w czasie lub zakres jego wartości. Z drugiej strony dane uzyskane z eksperymentu mogą wpływać na przyjęcie odpowiedniej struktury modelu $\mathcal{M}$. Bywają jednak też takie sytuacje, gdy powtórzenie eksperymentu jest praktycznie niemożliwe i wtedy co najwyżej powtarza się trzy ostatnie kroki.

Zwykle przy kolejnych nawrotach w procedurze występuje konieczność „poprawiania" $\mathcal{M}$ lub **R**. W zasadzie **R** i $\mathcal{M}$ powinny być określone *a priori*, przy wykorzystaniu informacji, która może być dostępna w postaci *faktów*, *założeń* lub *hipotez* oraz przy uwzględnieniu celu, dla jakiego model jest tworzony. Podstawowa różnica między *założeniami* a *hipotezami* polega na tym, że te ostatnie mogą być testowane na danych, podczas gdy założenia nie. Ponieważ błędne założenia mogą prowadzić do błędnych wyników bez możliwości sprawdzenia tego na etapie identyfikacji, stąd należy dążyć do tego, by było jak najmniej *założeń*, a jak najwięcej *hipotez*. Oznacza to, że powinno się *testować* te wszystkie wyniki modelowania, które mogą być

testowane. Niestety, niektórych założeń nie da się w ogóle testować. Jeśli więc takie założenia okażą się błędne, to będą one stanowiły swego rodzaju źródła „pułapek", por. 5.2.2. Oczywiście fakty jako takie nie mogą być błędne z samej definicji – fakty, które byłyby błędne, stanowiłyby założenia.

Istota iteracyjnej procedury identyfikacji polega na tym, że $\mathcal{M}$ oraz **R** niekoniecznie muszą być oparte jedynie na *faktach* i *założeniach*, ale mogą stanowić sekwencje *próbnych* struktur i *próbnych* eksperymentów, które w ostatnim kroku będą testowane jako *hipotezy*. Dlatego też procedura wymaga określenia *a priori* ciągu struktur hierarchicznych $\mathcal{M}(1)$, $\mathcal{M}(2)$, ..., $\mathcal{M}(m_{max})$, tj. takich, że struktury modeli $\mathcal{M}(m) = (\mathcal{F}, \mu_m)$ z większymi wartościami m są zawsze bardziej ogólne niż z mniejszymi, por. 4.1.1, oraz ciągu hierarchicznych eksperymentów $\mathbf{R}(1)$, $\mathbf{R}(2)$, ..., $\mathbf{R}(r_{max})$, gdzie eksperyment $\mathbf{R}(r)$ dla większych wartości r zawsze dostarcza więcej informacji (r może być związane np. z większą liczbą danych N albo lepszym pobudzeniem). Warto tu zauważyć, że parametr r wpływa na postać zbioru modeli niesfalsyfikowanych $\mathcal{N}(r)$, jako że rozszerzanie eksperymentu (większe r) powoduje zawężanie $\mathcal{N}$, bo mniej modeli przejdzie test falsyfikacji (5.3).

Zasadniczo sekwencje te określa się w sposób arbitralny. Najogólniejszą strukturą modelu jest $\mathcal{M}_{max} = \mathcal{M}(m_{max})$, a eksperyment dostarczający najwięcej informacji określony jest przez $\mathbf{R}_{max} = \mathbf{R}(r_{max})$. Ciąg $\{\mathcal{M}(m)\}$ umożliwia także rozróżnienie między założeniami a hipotezami. To, co może podlegać falsyfikacji, to liczby złożoności m, które reprezentują *hipotezy*, natomiast $\mathcal{F}$ występujące w $\mathcal{M}$ reprezentuje postać modelu opartą na *założeniach* i *faktach*.

Omówione w poprzednim punkcie zbiory modeli niesfalsyfikowanych $\mathcal{N}$ i modeli odpowiednich do celu $\mathcal{Q}$ można wykorzystać do zatrzymywania iteracyjnej procedury identyfikacji. Przyjmuje się, że $\mathcal{M}$, $\mathcal{N}$ oraz $\mathcal{Q}$ są podzbiorami większego zbioru zawierającego *wszystkie* modele M. Jeżeli w zbiorze modeli próbnych $\mathcal{M}$ wystąpią elementy wspólne ze zbiorami $\mathcal{N}$ i $\mathcal{Q}$, to procedura identyfikacji powinna się zakończyć sukcesem. Można także oczekiwać wystąpienia takich sytuacji, gdy model ze zbioru $\mathcal{N}$ nie będzie *odpowiedni do celu*, albo model ze zbioru $\mathcal{Q}$ zostanie *sfalsyfikowany*. Innymi słowy, model nie musi się zgadzać ze wszystkimi dostępnymi danymi, aby być wystarczająco dobry do założonego celu, i odwrotnie, może się dobrze zgadzać z danymi, ale być bezużyteczny.

Realizacja iteracyjnej procedury identyfikacji zależy od tego, jakie są *możliwości walidacji* oraz *falsyfikacji*, a więc od tego, czy istnieją odpowiednie testy umożliwiające zatrzymanie tej procedury. W praktyce zawsze jakiś model da się otrzymać. Jeśli jednak w ramach zadania identyfikacji istnieje możliwość sprawdzenia, czy model ten jest także odpowiedni do postawionego celu, to jednocześnie można potwierdzić jego przydatność jeszcze przed

zastosowaniem. Jeśli takiej możliwości nie ma, to model może być dobry, ale tego nie da się stwierdzić na etapie identyfikacji. Jeśli cel tworzenia modelu nie jest wyraźnie zdefiniowany lub brak jest możliwości sprawdzenia odpowiedniości modelu do postawionego celu, ale istnieje możliwość falsyfikacji modelu (co na ogół jest łatwiejsze), to jedyne co pozostaje, to szukanie najprostszego modelu, który nie będzie sprzeczny z danymi doświadczalnymi. Bohlin w [17] rozróżnia dwie sytuacje, dla których formułuje dwa różne podejścia:

1) Jeżeli brak jest jasno sprecyzowanego celu Q, co pociąga za sobą brak możliwości walidacji, to w tym przypadku proponuje on oparcie iteracyjnej procedury identyfikacji na tzw. *regule naukowej*, która jest prostą konsekwencją zastosowania ogólnej zasady poprawiania i falsyfikacji hipotez do zadania identyfikacji. Warunkami wstępnymi dla procedury jest określenie ciągów $\{\mathcal{M}(m)\}$ i $\{\mathbf{R}(r)\}$ oraz $m_{\max}$ i $r_{\max}$. Warunkami ograniczającymi dla procedury są albo przeprowadzenie eksperymentu o największej dopuszczalnej złożoności $r_{\max}$, albo osiągnięcie najogólniejszej dopuszczalnej struktury $m_{\max}$. W procedurze tej występują dwie pętle: wewnętrzna, w której następuje poprawianie eksperymentu, i pętla zewnętrzna, obejmująca poprawianie struktury. Wykonywanie pętli wewnętrznej kończy się, gdy model został *sfalsyfikowany*, co oznacza, że należy poprawić strukturę modelu, albo gdy osiągnięto ograniczenie $r_{\max}$. Jeżeli mimo wykonania najobszerniejszego eksperymentu model jest nadal niesfalsyfikowany, to kończona jest także pętla zewnętrzna i procedura kończy się sukcesem. Przedstawiona procedura została później zmodyfikowana dla celów identyfikacji nieliniowego modelu typu „szara skrzynka" dla obiektu stochastycznego, por. [18].

2) Jeśli jednak cel, któremu model ma służyć, jest wyraźnie określony, to wtedy można zakończyć pętle zewnętrzną i wewnętrzną wtedy, gdy osiągnięto model *odpowiedni do celu*. W takim przypadku procedura identyfikacji realizowana jest według *reguły inżynierskiej*, która stanowi logiczną konsekwencję uzupełnienia *reguły naukowej* o wprowadzony do niej *warunek walidacji* modelu. Pętla wewnętrzna kończy się wtedy, gdy uzyskano model *odpowiedni do celu*, co jest równoznaczne z zakończeniem także pętli zewnętrznej. Pętla wewnętrzna może być także zakończona wcześniej, zanim osiągnięto model *odpowiedni do celu* – wtedy, gdy model został *sfalsyfikowany* albo nie ma już możliwości poprawiania eksperymentu. W takiej sytuacji należy poprawić strukturę modelu i kontynuować procedurę. Procedura kończy się, jeśli otrzymano model *odpowiedni do celu* lub wyczerpano możliwości poprawiania struktury przez osiągnięcie $m_{\max}$.

Podejście takie, chociaż logicznie poprawne i *teoretycznie* uzasadnione przez metodologię badań naukowych, jest jednak w praktyce trudne do akceptacji, może poza wyjątkowymi

przypadkami występującymi w warunkach czysto laboratoryjnych. Trudno bowiem wyobrazić sobie w praktyce sytuację, aby można było zakładać wielokrotne powtarzanie eksperymentów w wewnętrznej pętli procedury. Ze względu na koszty związane z powtarzaniem eksperymentów powinno się raczej zaczynać od razu od najobszerniejszego eksperymentu $\mathbf{R}(r_{\max})$, jaki da się przeprowadzić, biorąc pod uwagę przyjęte założenia o najogólniejszej strukturze modelu $\mathcal{M}(m_{\max})$ oraz realne możliwości. Do ich określenia można wykorzystać inne źródła informacji wymienione w 5.2.3. Następnie powinno się testować w pętli kolejne struktury, poczynając od najprostszej, aż uzyskany zostanie model *odpowiedni do celu* i jednocześnie *niesfalsyfikowany*. Dopiero wtedy, gdy żadna ze struktur nie spełni postawionych wymagań, należy przemyśleć możliwość poszerzenia struktury $\mathcal{M}$, np. zwiększając zakres poszukiwania $m_{\max}$ lub zmieniając postać modelu $\mathcal{F}$ poprzez wzbogacenie jej np. o elementy nieliniowe albo zmienne w czasie, i dopiero wtedy powtarzać eksperyment (najobszerniejszy jaki się da) dla nowej hipotezy (tj. nowego ciągu struktur), o ile poprzedni eksperyment nie byłby już dla niej wystarczający. Być może przy takim postępowaniu niepotrzebnie zaczyna się od zbyt obszernego (a więc również kosztownego) eksperymentu, ale przynajmniej nie ma potrzeby wielokrotnego powtarzania kolejnych eksperymentów dla prostych struktur, które być może i tak nie spełnią pokładanej w nich nadziei. Praktyka identyfikacji pokazuje, że *per saldo* wyjdzie to i taniej, i szybciej.

5.2.1. Postać sformalizowana procedury

Po uwzględnieniu przedstawionych wyżej uwag zaproponowano iteracyjną procedurę identyfikacji w ujęciu praktycznym, którą można przedstawić w postaci bardziej formalnej wykorzystując konwencję strukturalnego języka programowania w następujący sposób (symbol lewej strzałki ← odpowiada tu instrukcji przypisania):

```
BEGIN (* Procedura identyfikacji w ujęciu praktycznym *)
   Określ F, m_min, m_max, R_max, V, Q(M, Δ, δ), Γ i γ (* Inicjalizacja pętli zewnętrznej *)
   REPEAT (* Pętla zewnętrzna *)
      R_max → (U_N , Y_N ) (* Wykonaj eksperyment *)
(* Początek części procedury, która może być wykonywana automatycznie przez komputer *)
      (* Inicjalizacja pętli wewnętrznej*)
      m ← m_min - 1; model niesfalsyfikowany; model nieodpowiedni do celu
      REPEAT (* Pętla wewnętrzna *)
         m ← m + 1 (* Popraw strukturę modelu *)
```

$\hat{M} \leftarrow \arg\min \{V(M \mid \mathbf{U}_N, \mathbf{Y}_N) \mid M \in \mathcal{M}(m)\}$ (* estymacja modelu *)

(* Czy model jest odpowiedni do celu? *)

IF $\Delta(\hat{M} \mid \mathbf{M}) \leq \delta$ THEN *model odpowiedni do celu*

ELSE *model nieodpowiedni do celu*

END_IF

(* Czy dane falsyfikują strukturę? *)

IF $\Gamma(\hat{M} \mid \mathbf{U}_N, \mathbf{Y}_N) > \gamma$ THEN *model sfalsyfikowany*

ELSE *model niesfalsyfikowany*

END_IF

(* Koniec pętli wewnętrznej *)

UNTIL (*model nieodpowiedni do celu* OR *model sfalsyfikowany*) AND ($m < m_{max}$)

(* Koniec części procedury, która może być wykonywana automatycznie przez komputer *)

IF (*model nieodpowiedni do celu* OR *model sfalsyfikowany*) THEN

Popraw $\mathcal{F}$ lub m_{max} lub R_{max}

END_IF

(* Koniec pętli zewnętrznej *)

UNTIL (*model nieodpowiedni do celu* OR *model sfalsyfikowany*)

END (* Koniec procedury identyfikacji w ujęciu praktycznym *)

W przedstawionej procedurze jedynie pętla wewnętrzna jest zapisana w sposób na tyle sformalizowany, że może być zaprogramowana w komputerze i wykonywana automatycznie, bez bezpośredniego udziału osoby wykonującej identyfikację. Ze względu na subiektywny charakter doboru czy poprawiania $\mathcal{F}$, m_{max} i R_{max}, które zależą od wiedzy *a priori*, celu identyfikacji oraz ewentualnie innych ograniczeń, a także od uzyskanych wyników zastosowanej procedury, trudno jest przedstawić je w sposób obiektywny i dlatego wykonywanie pętli zewnętrznej pozostaje w gestii użytkownika. W punkcie 5.2.4 przedstawiono kilka testów pomocniczych, które na podstawie danych $(\mathbf{U}_N, \mathbf{Y}_N)$ mogą ułatwić wybór postaci $\mathcal{F}$, natomiast w punkcie 5.3 omówiono szczegóły dotyczące realizacji pętli wewnętrznej w przypadku identyfikacji liniowych modeli stacjonarnych (*LTI*).

Teoretycznie rzecz biorąc można sobie wyobrazić również takie rozwiązanie, w którym zakłada się nie jedną postać $\mathcal{F}$, ale zdefiniowany *a priori* ciąg $\mathcal{F}_1, \mathcal{F}_2, \ldots, \mathcal{F}_{max}$, przy czym postać $\mathcal{F}$ o mniejszym indeksie jest prostsza od $\mathcal{F}$ o indeksie większym. Na przykład, dla liniowego modelu obiektu $\mathcal{F}_1$ może reprezentować model typu *FIR*, $\mathcal{F}_2$ model *ARX*, a $\mathcal{F}_{max}$ model *ogólny*, por. 4.2.1. Wtedy wykonywana automatycznie przez komputer pętla wewnętrzna mogłaby być zorganizowana w postaci dwóch pętli: jednej (zewnętrznej), w której poszukuje się $\mathcal{F}$,

poczynając od $\mathcal{F}_1$, oraz drugiej (wewnętrznej) dla poszukiwania μ. Z tym wiązałaby się jednak konieczność założenia różnych wartości m_{min} i m_{max} dla różnych $\mathcal{F}$ (np. dla modeli *FIR* m_{max} musiałoby być znacznie większe niż dla *ARX*) oraz zaproponowania takich testów, które umożliwiałyby porównywanie różnych $\mathcal{F}$. Kolejnym problemem mogłoby być określenie spójnych kryteriów falsyfikacji Γ, jako że dla różnych form $\mathcal{F}$ mogłyby być odpowiednie inne testy. Na dodatek dla różnych $\mathcal{F}$ mogą być określone inne wymagania na eksperyment R_{max} (np. inne są warunki *PE* dla różnych typów modeli, por. 5.4.2). Poza tym praktyczna wartość takiego rozwiązania byłaby wątpliwa, ponieważ to na osobie prowadzącej identyfikację powinna spoczywać odpowiedzialność za wybór $\mathcal{F}$, gdyż to ona ma określony cel identyfikacji oraz dodatkową wiedzę *a priori*, którą trudno byłoby w sposób formalny zapisać w procedurze. W 4.2.2 przedstawiono zestaw porad dotyczących wyboru $\mathcal{F}$ w przypadku modelu liniowego, które można przedstawić użytkownikowi w systemie komputerowego wspomagania identyfikacji.

Warunkami początkowymi w przedstawionej procedurze są określone przez użytkownika: postać modelu $\mathcal{F}$, zakresy m_{min} i m_{max} dla poszukiwania zbioru wskaźników μ_m i kryterium aproksymacji $\mathcal{A}$ zdefiniowane przez funkcję strat V. Przyjęto tu, że na podstawie wiedzy *a priori* można także ograniczyć zakres poszukiwania modelu od dołu przyjmując m_{min} większe od 0.

Oprócz tego, jako warunki początkowe należy określić cel Q, tzn. model $\mathbf{M}$ równoważny obiektowi $\mathbf{S}$ ze względu na cel identyfikacji, kryterium walidacji Δ i dokładność δ odpowiednio do (5.6), a także kryterium falsyfikacji Γ i poziom tolerancji γ zgodnie z (5.3). Dobór obu kryteriów nie jest sprawą prostą, zwłaszcza gdy chodzi o osiągnięcie pewnego kompromisu między wymaganą dokładnością modelu wynikającą z kryterium walidacji a odpornością na falsyfikację. Ze względu na potencjalne źródła błędów, por. 5.2.2, nie powinno się rezygnować z kryterium falsyfikacji, ale przyjęcie zbyt ostrego kryterium może z kolei prowadzić do konieczności kontynuowania identyfikacji mimo otrzymania już modelu *odpowiedniego do celu*. W ten sposób można osiągnąć model dokładniejszy niż potrzeba, co pociąga za sobą niepotrzebne koszty. W punkcie 5.4 omówiono propozycje różnych testów dla liniowych modeli parametrycznych, które mogą być wykorzystane jako kryteria dla określenia, czy model jest odpowiedni do celu albo niesfalsyfikowany.

Uzyskanie modelu odpowiedniego do celu i jednocześnie niesfalsyfikowanego kończy procedurę identyfikacji. W przeciwnym przypadku należy zwiększyć zakres poszukiwania m_{max} lub założyć bardziej złożoną postać modelu $\mathcal{F}$. Obie modyfikacje mogą się wiązać z koniecznością zaplanowania nowego eksperymentu R_{max}. Pewnym wyjściem z sytuacji może być czasem także zmniejszenie wymagań na dokładność modelu δ w kryterium walidacji lub *granicznej* wartości γ dla kryterium falsyfikacji.

Osobnym zagadnieniem jest problem, który tu został nazwany „poprawianiem struktury". Otóż jeżeli ciąg struktur hierarchicznych $\{\mathcal{M}(m)\}$ jest sekwencją zawierającą jedynie *proste struktury zagnieżdżone*, tj. takie, dla których każdemu m odpowiada tylko pojedynczy wskaźnik struktury μ_m, por. 4.1.1, to zadanie poprawiania struktury jest zdefiniowane w sposób jednoznaczny. Najprostszy przykład takiej sytuacji występuje wtedy, gdy μ jest skalarem. Inna możliwość to przyjęcie założenia, że każdy element wektora μ ma tę samą wartość. W takiej sytuacji problem sprowadza się do określenia rzędu modelu (ang. *model order determination*). Problem staje się poważny wtedy, gdy μ jest wektorem o większych wymiarach, a założenie prostych struktur zagnieżdżonych byłoby zbyt dużym ograniczeniem dla poszukiwania odpowiedniej struktury modelu. W takiej sytuacji mówi się o problemie określenia struktury modelu (ang. *model structure determination*), który ze względu na liczbę stopni swobody jest zagadnieniem znacznie trudniejszym. Algorytm poszukiwania odpowiedniej struktury dla liniowego modelu parametrycznego zaproponowano w punkcie 5.3.2.

Jeżeli brak jest możliwości ustalenia reguły dla *walidacji*, to pomija się testowanie warunku $\Delta(\hat{M}|\mathbf{M}) \leq \delta$ i przyjmuje, że model *niesfalsyfikowany* jest jednocześnie modelem *odpowiednim do celu*. Teoretycznie rzecz biorąc, warunek wystarczający na otrzymanie modelu *odpowiedniego do celu* jest oczywisty: $\mathcal{M} \cap \mathcal{N} \subseteq Q$ (tzn. w zbiorze testowanych struktur musi się znajdować model niesfalsyfikowany, który jednocześnie będzie odpowiedni do celu). Nie jest to jednak warunek konieczny, ponieważ model *odpowiedni do celu* wcale nie musi występować w zbiorze $\mathcal{M} \cap \mathcal{N}$. Poza tym warunku tego nie da się zastosować bezpośrednio, ponieważ zbiór Q zależy od nieznanego obiektu i nie może być wyznaczony w sposób bezpośredni. Tak więc bez przyjętego kryterium walidacji wynikiem procedury może być albo brak modelu (wszystkie próbne modele zostały sfalsyfikowane), albo uzyskano model, który nie został sfalsyfikowany. Jednak w tym drugim przypadku nie da się stwierdzić przed zastosowaniem modelu, czy jest on *odpowiedni do celu*, tj. że $\hat{M} \in Q$. Wynikiem identyfikacji może być bowiem taki model $\hat{M}$, który nie jest odpowiedni do postawionego celu, mimo że testy wykonane dla danych doświadczalnych nie powodują jego odrzucenia.

Jednak nawet wtedy, gdy nie określono kryterium walidacji, to istnieją pewne przesłanki na potwierdzenie tego, że jednak $\hat{M} \in Q$, por. [17]. Jest to możliwe wtedy, gdy przyjęto odpowiednio ogólną strukturę $\mathcal{M}(m_{\max})$ i wykonano taki eksperyment $\mathbf{R}_{\max}$, który zapewni, że model *niesfalsyfikowany* jest jednocześnie *odpowiedni do celu*. Istota takiego podejścia polega na nałożeniu tak silnych wymagań na eksperyment (np. długość trwania lub intensywność pobudzania), by tylko nieliczne modele przeszły testy falsyfikacji, tj. te, które na pewno znajdują się w zbiorze Q. Pętla w procedurze identyfikacji nie jest zatrzymywana wtedy, gdy

osiągnięto model odpowiedni do celu, ale wtedy, gdy struktura jest na tyle bogata, że otrzymany model przejdzie test falsyfikacji dla tak obszernego eksperymentu. Takie rozwiązanie, jeśli zakończy się sukcesem, może doprowadzić do modelu, który może być nawet zbyt dokładny z punktu widzenia celu, chociaż czasem może to być jedyne rozwiązanie. Podobne rozumowanie można przeprowadzić wtedy, gdy w trakcie eksperymentu pobudzenie obiektu miało takie same własności, jak pobudzenie występujące w czasie stosowania modelu, a celem identyfikacji jest uzyskanie modelu, którego odpowiedź naśladuje odpowiedź obiektu. W takim przypadku zbiór modeli *niesfalsyfikowanych* $\mathcal{N}$ pokrywa się ze zbiorem Q modeli *odpowiednich do celu*, a więc jest spełniony warunek $\mathcal{M} \cap \mathcal{N} \subseteq Q$.

5.2.2. Potencjalne źródła błędów

Przedstawiona wyżej ogólna procedura identyfikacji nie gwarantuje otrzymania poprawnego wyniku w każdej sytuacji. Zdarzają się bowiem takie przypadki, gdy identyfikacja kończy się uzyskaniem błędnego modelu, a jednocześnie nie można tego stwierdzić na podstawie zastosowanej procedury. To, że model jest błędny, okazuje się dopiero w czasie jego stosowania, z mniejszymi lub większymi tego konsekwencjami. Przypadek taki można określić jako *pułapkę*. Stanowi ona najgorszy możliwy wynik identyfikacji - na podstawie przeprowadzonej procedury identyfikacji wydaje się, że otrzymano dobry model, a tymczasem jest on błędny.

Przyjmując w domyśle, że teoria identyfikacji oraz zastosowane oprogramowanie są poprawne, a także, że niefortunny wynik identyfikacji nie jest efektem jakiegoś wyjątkowego zbiegu okoliczności, można powiedzieć, że pułapki wynikają głównie z błędnych *założeń* oraz z *niedostatecznie przetestowanych hipotez* [16].

Przyjęcie *założenia* oznacza, że twórca modelu uznaje, iż jakieś *warunki* są spełnione dla rzeczywistego systemu, obejmującego eksperymentatora **R** i obiekt **S**. Jeśli jednak warunki te nie będą spełnione, to w efekcie może to prowadzić do powstania różnego rodzaju błędów. Przykładami takich założeń są: *dostateczne pobudzenie*, *powtarzalność* czy *rozdzielność eksperymentatora od obiektu* (por. punkt 3.1). Ponieważ założenia te obejmują **R** i **S**, a w szczególności dotyczą możliwych ukrytych zależności wzajemnych, więc mogą stanowić szczególnie dobrze ukrytą przyczynę powstawania błędów. Przykładem takiego błędu mogłaby być próba identyfikacji obiektu elektroakustycznego dla celów aktywnego tłumienia hałasu w trakcie pracy układu kompensacji, co omówiono bliżej w punkcie 2.3.2. W takim przypadku bowiem niezauważenie jednoczesnej zależności sygnałów wyjściowego oraz pobudzającego od trzeciej zmiennej, tj. sygnału hałasu, nie umożliwia poprawnej identyfikacji toru wtórnego.

Podstawowy problem, jaki się tu pojawia, to odróżnienie *korelacji* (zależności statystycznej) od *przyczynowości*, por. np. [17]. To, co daje się zaobserwować jako wynik eksperymentu, zawsze stanowi tylko korelację – **u** i **y** zmieniają się w jakimś szczególnym wzajemnym powiązaniu. Tymczasem dla celów zastosowania wymagane jest istnienie *przyczynowości*, czyli stwierdzenie, że **u** powoduje zmianę **y** w sposób, który można opisać za pomocą modelu. Przyjęcie modelu opartego na doświadczeniu i zaakceptowanie go do zastosowania jest na dobrą sprawę efektem pewnego *założenia*, że obserwowana korelacja między **u** a **y** jest rzeczywiście spowodowana przez związek przyczynowo-skutkowy. Jeżeli korelacja taka *nie jest* efektem zależności, ale spowodowana jest przez *ukrytą wspólną zmienną*, to należy oczekiwać, że wynik identyfikacji będzie błędny, chociaż sam model może podawać prawidłowe wyniki tak długo, jak długo pozostaje w mocy zależność między trzema rozpatrywanymi zmiennymi. Zależność statystyczna sama z siebie nigdy *nie może być dowodem* zależności przyczynowo-skutkowej. Uzasadnienie stwierdzenia takiej zależności musi pochodzić *spoza* statystyki, np. musi wynikać z *dodatkowej wiedzy* o procesie, która może pochodzić z analizy teoretycznej lub z innych źródeł. Sytuacja na ogół jest jasna w przypadku obiektów technicznych – określenie, co jest *wejściem* (przyczyną), a co *wyjściem* (skutkiem), wynika bezpośrednio z zasady działania obiektu. Jednak i tu mogą się pojawić dodatkowe wpływy zmiennych ukrytych.

Przyjęcie *hipotezy* z kolei związane jest z koniecznością wykonania odpowiednich *testów*, które umożliwią odrzucenie jej, gdy jest ona błędna. Najczęściej hipotezy reprezentowane są przez zbiór wskaźników struktury μ. Niewykonanie więc odpowiednich testów falsyfikujących μ, np. na drodze testowania funkcji autokorelacji reszt, może prowadzić do przyjęcia zbyt prostego modelu. Ale nawet wykonanie takich testów może nie wystarczyć, bowiem samo badanie autokorelacji bez badania wyższych momentów nie umożliwia wykrycia nieliniowości w identyfikowanym procesie. Może się więc okazać, że potrzebne jest także wykonanie odpowiednich testów falsyfikujących całą strukturę $\mathcal{M}$.

Inne możliwości powstania potencjalnych błędów w trakcie identyfikacji związane są z kryterium walidacji. Podstawowym źródłem błędu może być tu wybór takiego kryterium walidacji, które nie jest adekwatne do celu, dla którego model jest tworzony, zwłaszcza że kryterium to często bywa określone w sposób heurystyczny albo wynika z ograniczeń wnoszonych przez oprogramowanie (np. program wymaga, aby była to funkcja błędów predykcji). W rozpatrywanym przykładzie identyfikacji obiektu elektroakustycznego do wyboru struktury modelu w procedurze identyfikacji stosowane będzie kryterium informacyjne, które jest funkcją błędów predykcji, co wynika z zastosowanego oprogramowania. W celu poprawnej pracy układu natomiast wymagane jest uzyskanie modelu o odpowiednio dokładnej

charakterystyce amplitudowo-fazowej. Stąd konieczność dodatkowego testowania tej charakterystyki, por. 5.1.4.

Aby móc sprawdzić, czy model $\hat{M}$ jest *odpowiedni do celu* Q, trzeba określić model **M** równoważny obiektowi ze względu na cel, por. 5.1.3. Tak więc błędne zdefiniowanie **M** może także prowadzić do błędnego wyniku identyfikacji. W omawianym przykładzie walidacji modelu obiektu elektroakustycznego za **M** przyjęto charakterystykę amplitudowo-fazową uzyskaną metodą nieparametryczną. Pojawia się więc niebezpieczeństwo polegające na tym, że w przypadku błędnego zidentyfikowania tej charakterystyki kryterium walidacji polegające na porównaniu modeli uzyskanych metodą parametryczną i nieparametryczną może prowadzić do błędnego wyboru modelu parametrycznego. Jednak z braku innych możliwości zdefiniowania modelu równoważnego obiektowi było to jedyne rozsądne podejście. Dlatego też przyjęto, że **M** należy wyznaczyć na podstawie dodatkowych eksperymentów, zaplanowanych pod kątem uzyskania odpowiedniej dokładności charakterystyki częstotliwościowej.

Z drugiej strony, jeśli nie określi się reguły walidacji, to zwiększa się możliwość pojawienia się zagrożeń, ponieważ wykonywanych jest mniej testów. To samo z siebie nie musi prowadzić do powstania pułapki w tym sensie, że jakieś fałszywe hipotezy zostaną przyjęte za prawdziwe i procedura zakończy się np. otrzymaniem błędnego μ. Problem polega tu na tym, że μ jako wynik zastosowania procedury zależy nie tylko od założonej maksymalnej złożoności modelu m_{max}, ale także od przeprowadzonego eksperymentu $\mathbf{R}_{max}$. I tu może się pojawić pułapka. Jeśli bowiem eksperyment ten będzie niewystarczający, aby uzyskać model *odpowiedni do celu*, np. w trakcie eksperymentu nie zastosowano pobudzenia o takich własnościach, jakie mogą wystąpić na etapie stosowania, por. 3.1, to brak jest przesłanek, które umożliwiłyby wykrycie takiej sytuacji. Dlatego też planując eksperyment dla identyfikacji obiektu elektroakustycznego przyjęto, że w sygnałach pobudzających odpowiednia wartość mocy powinna występować dla wszystkich linii widma do częstotliwości Nyquista, por. 3.4.1. Podobnie, nawet jeśli struktura modelu przejdzie testy falsyfikacji, to wcale nie musi być odpowiednia do danego celu. Na przykład, spełnienie przez model testu białości reszt nie oznacza, że jego charakterystyka amplitudowa czy fazowa są odpowiednio dokładne.

Natomiast wybór modelu $\hat{M}$ związany z optymalizacją określonego kryterium aproksymacji $\mathcal{A}$ (np. minimalizacją sumy kwadratów błędów predykcji) sam w sobie nie stanowi źródła pułapek, bowiem w przypadku osiągnięcia w metodzie estymacji parametrów minimum lokalnego lub przedwczesnego zatrzymania procedury minimalizacji zazwyczaj kończy się to odrzuceniem modelu.

Podsumowując, jeśli w procedurze identyfikacji nie można skorzystać z odpowiedniego kryterium walidacji, to wynik jest bardziej zależny od sposobu przeprowadzenia eksperymentu

oraz od informacji *a priori*. Wymagania teoretyczne wystarczające do wykonania eksperymentu w sposób właściwy omówiono w punkcie 3.1, jednak w praktyce może się pojawić wiele przeszkód, które uniemożliwią poprawne przeprowadzenie doświadczenia i stąd ryzyko zakończenia procedury identyfikacji błędnym wynikiem. Jeśli natomiast w celu przyjęcia odpowiedniej struktury modelu czy poprawnego zaplanowania eksperymentu chce się skorzystać z innych informacji niż dane, to przy korzystaniu z takich informacji trzeba zachować odpowiednią ostrożność. W takiej sytuacji bowiem mamy do czynienia z większą liczbą założeń niż hipotez, jako że mniej informacji może być testowanych.

Warto tu zauważyć, że przedstawione wyżej możliwe przyczyny powstania błędnego wyniku identyfikacji związane są głównie z tymi elementami w procedurze identyfikacji, które nie mogą być w sposób automatyczny zaprogramowane w procedurze, a więc za które odpowiada użytkownik. Tego bowiem zazwyczaj nie da się wydedukować operując tylko na danych i korzystając z ogólnych teorii na temat wnioskowania.

5.2.3. Inne źródła informacji

Jak już stwierdzono wyżej, same dane doświadczalne $(\mathbf{U}_N, \mathbf{Y}_N)$ nie stanowią wystarczającej informacji, aby uzyskać prawidłowy wynik identyfikacji procesu. Potrzebna jest jeszcze dodatkowa wiedza dla określenia $\mathcal{M}$, **R** oraz wyznaczenia $\mathcal{N}$. W trakcie prowadzenia identyfikacji przydatne mogą być więc także inne dostępne źródła informacji, por. np. [17].

Do źródeł tych należy przede wszystkim *informacja a priori* o obiekcie **S**, wynikająca np. z praw fizyki, poprzednio otrzymanych zależności empirycznych, wartości parametrów mierzonych niezależnie, a nawet z subiektywnych opinii czy przypuszczeń. Ocena rzetelności takich *założeń a priori* może także stanowić użyteczną informację. Tego typu wiedzę można wykorzystać do ograniczenia zbioru modeli próbnych, tj. struktury $\mathcal{M}$. Podobnym źródłem informacji mogą być dane pochodzące z *eksperymentów a priori*, tj. eksperymentów prowadzonych wcześniej po to, aby poprawnie zaprojektować właściwy eksperyment identyfikacyjny. Na przykład, wybór częstotliwości próbkowania powinien zależeć od dominującej stałej czasowej w obiekcie, a więc przy braku tego typu wiedzy konieczne może być wstępne pobudzenie obiektu sygnałem skokowym w celu jej oszacowania, por. 3.2.3. Eksperymenty *a priori* można wykorzystać także do określenia zbioru $\mathcal{N}$ modeli *niesfalsyfikowanych*, np. jeżeli z odpowiedzi skokowej wynika, że obiekt ma własności oscylacyjne, to wiadomo, że model także powinien tego typu własności przejawiać, por. testy wizualizacyjne dla modeli liniowych omówione w 5.4.1.

Informacja a priori jest tak samo istotna jak dane doświadczalne $(\mathbf{U}_N, \mathbf{Y}_N)$. To, że w wielu praktycznych sytuacjach istnieje tendencja do polegania bardziej na danych rzeczywistych, pochodzących z doświadczenia na obiekcie, niż na przyjętych z góry założeniach, wynika stąd, iż brak jest pewności co do prawdziwości informacji *a priori*. Jednak informacja taka może być oparta także na innych danych doświadczalnych, równie pewnych (czy niepewnych) jak zarejestrowana odpowiedź **y** na pobudzenie **u**, które same w sobie też mogą być zakłócone.

Jeżeli po wykonaniu identyfikacji można przeprowadzić na rzeczywistym obiekcie **S** doświadczenie *a posteriori* w celu potwierdzenia przydatności modelu, to takie badanie jest oczywiście decydujące. Jeśli jednak nie można eksperymentować na obiekcie, lecz tylko symulować uzyskany model, to taki eksperyment *a posteriori* nie niesie już żadnej dodatkowej informacji, której nie dałoby się wywnioskować na podstawie danych. Z drugiej strony badania symulacyjne modelu mogą stanowić istotną pomoc przy jego ocenie, por. 5.4.1.

Wybór struktury $\mathcal{M}$ i eksperymentatora **R** nie stanowi zadań czysto matematycznych, które nie mogą być rozwiązane na drodze zastosowania jakiejś uniwersalnej teorii do zebranych danych. Stąd tak duże znaczenie ma tu także nabyte *doświadczenie* w identyfikacji procesów, tzw. „wyczucie inżynierskie". Wyczucie to może np. pomóc w określeniu, jak należy planować eksperyment, jak modelować obiekt o pewnych własnościach albo też jaki sposób postępowania może prowadzić do błędnych wyników i dlaczego. Wszystko to stanowi *zewnętrzne* źródło informacji, które można wykorzystać do przyjęcia założeń na temat struktury modelu $\mathcal{M}$ oraz planowania eksperymentu **R**. Jednak w przeciwieństwie do źródła informacji *a priori* o obiekcie **S** „wyczucie inżynierskie" jest niezależne od szczególnego przypadku modelowanego obiektu i powinno dać się opisać w sposób bardziej ogólny. Pewne usiłowania w tym kierunku można znaleźć w [16]. W pracy częściowo zagadnienia te przedstawiono w punktach 3.2 i 3.3, gdzie omówiono praktyczne możliwości wspomagania w zakresie planowania eksperymentu i przygotowania danych, oraz w 4.2.2 przy okazji dyskusji wyboru typu $\mathcal{F}$ modelu liniowego i w 5.2.4, gdzie przedstawiono testy wspomagające wybór struktury. Zalecenia czy zasady, które powinny być wzięte pod uwagę przy wyborze $\mathcal{M}$ czy **R**, są zazwyczaj formułowane w postaci intuicyjnej czy opisowej – występują więc w systemie komputerowego wspomagania w postaci porad. Dzieje się tak nie tyle dlatego, że tego w ogóle nie da się zrobić w postaci bardziej precyzyjnej, ale raczej, że jest to zbyt trudne ze względu na wiele możliwości występujących w praktyce.

5.2.4. Testy wspomagające wybór struktury

W przedstawionej wyżej procedurze identyfikacji użytkownik powinien określić postać $\mathcal{F}$ modelu na podstawie wiedzy *a priori* oraz celu identyfikacji. Jednak do systemu komputerowego wspomagania identyfikacji można wprowadzić także pewne proste testy, które mogą ułatwić użytkownikowi wykonanie zadania modelowania, a przede wszystkim pozwolą na:

- wstępną ocenę zmienności własności dynamicznych procesu w czasie,
- wstępną ocenę nieliniowości procesu,
- wstępną ocenę struktury modelu ciągu czasowego.

Czasem już na podstawie prostej oceny wzrokowej przebiegu danych da się zauważyć, że własności dynamiczne procesu zmieniają się w czasie. Jednak ocena taka wymaga pewnego doświadczenia, a poza tym opiera się ona na subiektywnym odczuciu. Stąd, aby wspomóc użytkownika w tym względzie, można zaproponować testy wstępnej oceny zmienności własności dynamicznych identyfikowanego procesu oparte na analizie przebiegu w czasie *współczynników* najprostszych modeli parametrycznych estymowanych przy pomocy *Rekurencyjnej Metody Najmniejszych Kwadratów* (*RLS*) dla arbitralnie przyjętego *współczynnika zapominania* $\alpha = 0.98$, por. [101]. Zakłada się tu, że jeżeli proces jest niestacjonarny i jego własności zmieniają się w czasie, to fakt ten powinien znaleźć odzwierciedlenie nawet dla najprostszego modelu o parametrach zmiennych w czasie.

Do wstępnej oceny nieliniowości procesu wykorzystuje się zwykle funkcje statystyczne wyższych rzędów (ang. *Higher Order Statistics*), takie jak np. kurtoza i skośność, por. [106].

Z kolei wykresy funkcji autokorelacji i korelacji cząstkowych mogą być wykorzystane do określenia *struktury modelu ciągu czasowego* (tj. *AR*, *MA* czy *ARMA*) oraz oszacowania *stopnia* odpowiedniego wielomianu opisującego ciąg, por. [19].

5.3. Wspomaganie w identyfikacji struktury modeli liniowych

Zgodnie z uwagami przedstawionymi w 4.2.2 oraz 5.2 przyjmuje się, że użytkownik systemu komputerowego wspomagania identyfikacji powinien samodzielnie dokonać wyboru typu $\mathcal{F}$ modelu liniowego na podstawie pewnej wiedzy *a priori* o procesie, celu identyfikacji oraz wskazówek oferowanych przez ten system. Natomiast zadanie określenia zbioru wskaźników struktury μ może zostać sformalizowane w taki sposób, że da się go zaprogramować w komputerze, aby wektor ten był wyznaczany przez system w sposób automatyczny.

Oczywiście, użytkownik może także podać go bezpośrednio, jeżeli dysponuje odpowiednią wiedzą *a priori*. W dalszym ciągu przez identyfikację struktury modelu rozumie się więc automatyczne wyznaczenie nieznanego μ dla wybranego przez użytkownika typu $\mathcal{F}$.

Ponieważ nie jest znany estymator umożliwiający bezpośrednie wyznaczenie μ (poza modelami *AR* i *MA*, por. [19]) i trudno wyobrazić sobie zaproponowanie takiego estymatora, stąd na ogół robi się to na drodze estymacji modeli dla kolejnych struktur z zadanego przez użytkownika zakresu zmian dla μ i poddawania ich odpowiednim testom (pętla wewnętrzna w ogólnej procedurze identyfikacji, por. 5.2.1). Problemem, jaki się tu pojawia, jest przede wszystkim określenie *kryterium wyboru* struktury modelu. Ponieważ w praktyce trudno jest podać *kryterium walidacji* modelu w takiej postaci, jak to przedstawiono w sposób ogólny w 5.1.3 (a więc trudno jest wykonywać pętlę wewnętrzną w procedurze identyfikacji, aż uzyskany zostanie model, który spełni to kryterium), stąd najczęściej jako *substytut* kryterium walidacji przyjmuje się pewien *wskaźnik jakości* Q bazujący na zdefiniowanym w jakiś sposób *błędzie identyfikacji*, który może być uzupełniony o testy dodatkowe. Użytkownik powinien mieć tu możliwość wyboru takiego wskaźnika, który jego zdaniem najlepiej pasuje do przyjętego celu identyfikacji.

Drugą trudność stanowi określenie ciągu struktur hierarchicznych $\{\mathcal{M}(m)\}$, por. 4.1.1, które powinny być testowane w ramach wykonywania pętli wewnętrznej. Jeżeli z góry nie wprowadza się ograniczeń typu, że każdy wielomian jest tego samego stopnia (co w przypadku modelu *MISO* byłoby trudno uzasadnić), to wtedy zwykle liczba możliwych struktur jest *zbyt duża*, aby można je było testować na drodze przeglądu zupełnego. W takim przypadku konieczne jest określenie *strategii poszukiwań*, która pozwoliłaby na znalezienie optymalnej (w sensie wybranego kryterium Q) struktury modelu w rozsądnej liczbie kroków. Ponieważ zazwyczaj kryterium to zależy od struktury modelu w sposób silnie nieliniowy, więc strategia taka niestety nie gwarantuje znalezienia modelu optymalnego globalnie.

5.3.1. Kryterium wyboru struktury

Zgodnie z zasadą największej wiarygodności *ML* estymacja parametrów modelu θ sprowadza się do minimalizacji funkcji strat dla założonej *a priori* struktury modelu $\mathcal{M} = (\mathcal{F}, \mu)$, por. 5.1.1. Jeżeli jednak przyjąć, że zadanie identyfikacji polega nie tylko na wyznaczeniu θ, ale również nieznanego μ, to wtedy minimalizacji podlega funkcja strat przedstawiona w postaci ogólnej jako:

$$V(\mu, \theta \mid \mathcal{F}, \mathbf{U}_N, \mathbf{Y}_N) = -\ln L(\mu, \theta \mid \mathcal{F}, \mathbf{U}_N, \mathbf{Y}_N)/N, \tag{5.9}$$

co w przypadku modelu liniowego (4.9) sprowadza się do minimalizacji sumy kwadratów błędów predykcji (4.11).

Powszechnie wiadomo jednak, że tak zdefiniowane kryterium aproksymacji nie może samo w sobie stanowić podstawy do wyboru struktury modelu, ponieważ w takim przypadku im bogatsza struktura, tym mniejszy błąd predykcji. Tak więc poszukiwanie zakończy się albo na przyjętym ograniczeniu górnym μ_{max}, albo z chwilą powstania braku możliwości rozwiązania zadania aproksymacji spowodowanego słabym uwarunkowaniem estymatora. Stąd albo trzeba określić kryterium wyboru struktury inne niż kryterium aproksymacji, albo podać takie kryterium porównawcze dla modeli, w którym oprócz kryterium aproksymacji uwzględnia się także pewną „karę" za złożoność modelu (liczbę parametrów). W literaturze można znaleźć wiele różnych sposobów rozwiązania tego zadania.

Pionierskie podejście do tego problemu oparte na teorii informacji przedstawił Akaike w postaci kryterium *AIC* (*Akaike Information Criterion*) [2]. Bazuje ono na *średniej* odległości informacyjnej Kullbacka-Leiblera [71] między prawdziwą a estymowaną gęstością prawdopodobieństwa (podejście probabilistyczne), przy założeniu że liczba danych N jest wystarczająca. Inną zaproponowaną metodą estymacji rzędu modelu jest kryterium minimalnej długości opisu *MDL* (*Minimum Description Length*) [118], które w przypadku dwustopniowej metody *MDL* [8] polega na wybraniu takiej oceny *ML*, która zapewnia minimalną długość kodu odpowiadającego definicji Shannona. Z kolei kryterium *BIC* (*Bayesian Information Criterion*) [125] zapewnia maksimum prawdopodobieństwa *a posteriori* dla przyjętego *a priori* prawdopodobieństwa porównywania zbioru modeli. Kryterium to jest podobne do dwustopniowej metody *MDL*. W [7] przedstawiono geometryczną interpretację powyższych kryteriów i zaproponowano modyfikację *MDL* dla krótkich ciągów danych.

W literaturze można także znaleźć wiele innych propozycji definiowania kryterium informacyjnego, mniej lub bardziej uzasadnionych teoretycznie, jak np. *CIC* (*Control IC*) [20], którego celem jest wybór struktury modelu dla stochastycznej regulacji adaptacyjnej, *EIC* (*Empirical IC*) [13] przeznaczone głównie dla celów prognozowania, *FIC* (*Focussed IC*) [21] nastawione na wybór struktury uwzględniającej jeden szczególnie interesujący parametr, *GIC* (*Geometric IC*) [55] wykorzystywane do wykrywania zniekształceń geometrycznych, *SIC* (*Subspace IC*) [134] oparte na analizie funkcjonalnej i inne. W [10] z kolei zaproponowano podejście bazujące na nowej mierze odległości definiowanej dla zbioru modeli, nazwane minimalną złożonością opisu *MDC* (*Minimum Description Complexity*). Odległość między prawdziwym modelem a elementami zbioru modeli jest tu definiowana jako *minimum* odległości informacyjnej Kullbacka-Leiblera. Wspomniane wyżej *AIC* oraz *MDL* stanowią szczególne przypadki *MDC*.

Przedstawione wyżej propozycje różnych kryteriów można podzielić na dwie grupy: takie, które estymują odległość (względną) wg Kullbacka-Leiblera i usiłują wybrać odpowiedni model opierając się na zasadzie oszczędności, oraz takie, których celem jest uzyskanie zgodnej oceny wymiaru prawdziwego modelu, przy założeniu, że model ten istnieje i znajduje się w zbiorze modeli testowanych. Kryteria pierwszego typu (np. *AIC*) nie zakładają istnienia prawdziwego modelu, ani tego, że znajduje się on w zbiorze rozpatrywanych modeli. Celem drugiego typu kryteriów (np. *BIC*) jest znalezienie tego prawdziwego modelu, co z kolei wymaga odpowiednio licznego zbioru danych N.

Podsumowując, w większości znanych z literatury kryteriów informacyjnych idea ich stosowania sprowadza się do wyboru takiego modelu, dla którego wartość kryterium

$$IC(\mu, \theta \mid \mathcal{F}, \mathbf{U}_N, \mathbf{Y}_N, \alpha,) = V(\mu, \theta \mid \mathcal{F}, \mathbf{U}_N, \mathbf{Y}_N) + \alpha(N)m(\mu) \tag{5.10}$$

jest minimalna, gdzie $m(\mu)$ jest liczbą parametrów modelu (wymiarem wektora θ) zależną od zbioru wskaźników struktury μ, a $\alpha(N)$ współczynnikiem służącym wprowadzeniu pewnej „kary" za złożoność modelu, zależnym od typu stosowanego kryterium. Oczywiście, można także wyobrazić sobie kryteria o innej postaci, np. $V(\mu, \theta \mid \mathcal{F}, \mathbf{U}_N, \mathbf{Y}_N)\,[1 + \alpha(N)m(\mu)]$.

Wyboru struktury modelu na drodze minimalizacji kryterium uwzględniającego „karę" za liczbę parametrów można także dokonać korzystając z kryterium *FPE* (*Final Prediction Error*) [1], a w przypadku gdy identyfikacja prowadzona jest dla celów regulacji, kryterium tym może być *FOE* (*Final Output Error*) [143].

Inne z kolei podejście do zagadnienia wyboru struktury modelu wynika z zastosowania teorii testowania hipotez *ANOVA* (ang. *ANalysis Of VAriance*), por. np. [22], szczególnie poprzez wykorzystanie *testu F* Snedecora, por. np. [79]. Natomiast dla metod identyfikacji w dziedzinie częstotliwości zaproponowano dwustopniową metodę estymacji rzędu modeli SISO [119].

Projektując komputerowy system wspomagania identyfikacji trzeba podjąć decyzję, jakie kryteria wyboru struktury modelu zastosować. Dla każdej z przedstawionych wyżej propozycji można pokazać szczególne przypadki, dla których dane kryterium daje najlepsze wyniki, tymczasem w systemie należałoby zaproponować takie, które byłyby najbardziej uniwersalne. Ponadto należy uwzględnić fakt, że użytkownik takiego systemu na ogół nie jest specjalistą w zakresie statystyki. Stąd dając mu możliwość decydowania, według jakiego kryterium system powinien poszukiwać struktury modelu, dobrze jest dać mu do wyboru takie kryteria, które mogą być dla niego intuicyjnie bardziej zrozumiałe. Dlatego też w systemie *MULTI-EDIP* jako kryterium wyboru struktury zaproponowano zastosowanie dwóch preferowanych w literaturze kryteriów informacyjnych:

$$AIC(m) = N \ln V(\mu, \theta) + 2m, \tag{5.11}$$

oraz

$$BIC(m) = N \ln V(\mu, \theta) + m \ln N. \tag{5.12}$$

W pierwszym z nich kara za liczbę parametrów jest najmniejsza spośród wszystkich kryteriów informacyjnych, w drugim największa. Pozwala to wyraźnie odróżnić dwie sytuacje, dzięki czemu użytkownik bazując na pewnej wiedzy *a priori* oraz uwzględniając cel tworzenia modelu może jasno określić, czy bardziej zależy mu na prostocie modelu (wtedy wybiera *BIC*), czy na dokładności predykcji wyjścia (wtedy wybiera *AIC*).

W systemie *MULTI-EDIP* zrezygnowano ze stosowania innych zasad wyboru struktury modelu, takich jak *test F* czy kryterium *FPE*, gdyż – jak pokazano w [129] – istnieje bliski związek między tymi kryteriami a *AIC*, co zostało także potwierdzone przez własne badania autora.

Oprócz kryteriów informacyjnych, użytkownik systemu powinien mieć także możliwość wykorzystania do wyboru struktury innego wskaźnika oceny jakości modelu niż ten, który był użyty w zadaniu aproksymacji. Dwie najprostsze możliwości, które także zaproponowano w systemie *MULTI-EDIP*, to:

- Wybór struktury na podstawie minimum wariancji błędów predykcji (funkcji strat) obliczonej dla innego zbioru danych niż wzięty do estymacji (tzw. zbioru danych do testowania).

  Opcja ta jest możliwa tylko wtedy, gdy zbiór taki jest osiągalny, a jej istota wynika z zaobserwowanego w praktyce zachowania się tego kryterium. Otóż przewymiarowanie struktury modelu (nadparametryzacja) skutkuje zwykle tym, że model staje się bardzo dobrze dopasowany do danych wziętych do estymacji (a nawet zbyt dobrze, gdyż usiłuje wyjaśnić to, co nie jest własnością procesu, lecz wynika z losowości danych), natomiast to samo kryterium aproksymacji zastosowane do innego zbioru danych wykazuje pogorszenie dopasowania w miarę powiększania struktury, por. np. [101].

- Wybór na podstawie wariancji błędów symulacji.

  Jeżeli zadanie polega na identyfikacji modelu do celów sterowania, to wybór struktury można oprzeć na kryterium błędu wyjścia *FOE* (*Final Output Error*), por. [143], określonego wzorem:

$$FOE(m) = \frac{N+m}{N-m}\left[y(i) - \sum_{k=1}^{q} \hat{G}_k\left(z^{-1}\right) u_k(i) \right]^2, \tag{5.13}$$

gdzie $\hat{G}_k\left(z^{-1}\right)$ jest modelem k-tego toru sterowania. Wybór struktury minimalizującej wartość *FOE* zapewnia odpowiednie modelowanie dynamiki torów sterowania. Z oczywistych względów kryterium to jest niedostępne w przypadku identyfikacji modeli ciągów czasowych.

W dalszej części zakłada się więc, że użytkownik określa kryterium wyboru struktury modelu Q spośród czterech dostępnych: *AIC*, *BIC*, wariancja błędów predykcji dla zbioru danych do testowania lub wariancja błędów symulacji.

5.3.2. Algorytm poszukiwania dla modeli MISO

Określenie algorytmu poszukiwania struktury modelu minimalizującej odpowiednie kryterium wyboru Q ma szczególne znaczenie w sytuacji, gdy brak jest informacji *a priori* umożliwiającej ograniczenie zakresu poszukiwania dla struktury modelu, jak np. w sytuacji identyfikacji obiektu elektroakustycznego, zwłaszcza wtedy, gdy wymiar wektora μ jest znaczny. W takim przypadku bowiem, ze względu na zbyt dużą liczbę stopni swobody, nie można tego zrobić korzystając z metody przeglądu zupełnego i trzeba zaproponować taką strategię poszukiwania, która umożliwiłaby osiągnięcie wyniku w rozsądnej liczbie kroków.

Literatura dotycząca opisu jakiejś strategii poszukiwania struktury modelu jest bardzo uboga. W przyborniku do identyfikacji *Identification Toolbox* dla środowiska *MATLAB* w przypadku identyfikacji modelu *SISO* użytkownik musi podać zakres zmienności poszczególnych parametrów struktury, a wybór struktury dokonywany jest na drodze przeglądu zupełnego, por. [81]. Rozwiązanie to, jeśli ma przynieść wynik w rozsądnym czasie, wymaga przyjęcia dość wąskiego zakresu zmienności wskaźników struktury, a i tu dla modelu typu *BJ* może stanowić pewien problem. Natomiast w przypadku modelu *MISO*, w celu uproszczenia zadania poszukiwania przyjęto, że wszystkie wielomiany w modelu są tego samego stopnia. O ile podejście takie może być jeszcze jakoś uzasadnione w sytuacji identyfikacji modeli *ARX* dla obiektu, w którym poszczególne tory sterowania cechują się bardzo podobną dynamiką (chociaż i tu trudno z góry zakładać, że dynamika w każdym z torów musi być tego samego rzędu), to w przypadku identyfikacji modelu *MISO* dla obiektu, w którym poszczególne tory sterowania różnią się istotnie między sobą (jak np. w modelu *BJ*), może to prowadzić do uzyskania struktury modelu nieadekwatnej do rzeczywistego procesu. W [127] przedstawiono opis systemu umożliwiającego automatyczny wybór struktury tylko dla modeli SISO. W [45] opisano system ekspercki dla identyfikacji, w którym zaimplementowano oparty na regułach heurystycznych algorytm poszukiwania struktury modelu *MISO*. Przeznaczony jest on jednak tylko dla specyficznej struktury *ARARX* (oraz *ARX*). Poza tym nadaje się do stosowania dla

typowych obiektów, których dynamikę można modelować w postaci filtrów dolnoprzepustowych niskiego rzędu.

Tu natomiast przedstawiono nowy algorytm poszukiwania struktury modelu dla obiektów *MISO*, który został zaimplementowany w systemie *MULTI-EDIP*. W dalszym ciągu zakłada się, że użytkownik systemu komputerowego wspomagania identyfikacji powinien określić maksymalne wartości ograniczające zakres poszukiwań od góry, zarówno dla stopni wielomianów n_{max}, jak i dla czasów opóźnienia d_{max}. Ponieważ użytkownik może także dysponować pewną wiedzą *a priori* na temat wartości minimalnych, przyjmuje się jednocześnie, że może on ograniczyć zakres poszukiwań od dołu deklarując odpowiednie wartości n_{min} i d_{min}. W przypadku braku jakiejkolwiek wiedzy dolne ograniczenia ustala się na wartości 1.

Poszukiwanie stopni wielomianów w modelu można przeprowadzić zaczynając od dołu lub od góry. Pierwsze podejście jest bardziej naturalne i jego zaletą jest to, że złożoność kolejnych testowanych modeli jest mniejsza niż modelu końcowego. Z kolei poszukiwanie struktury modelu od góry pozwala z większym prawdopodobieństwem uniknąć lokalnych minimów, por. [127]. Jednak wadą tego rozwiązania jest możliwość powstania problemów numerycznych przy identyfikacji modeli nadparametryzowanych oraz problemy związane z redukcją rzędu modelu. Zaproponowana tam metoda redukcji na drodze wyznaczania par zero/biegun, które można uprościć, jest mało dokładna, ponieważ, jak pokazują badania symulacyjne, obliczone pierwiastki wielomianów na ogół charakteryzują się małą dokładnością (szczególnie jeśli chodzi o licznik transmitancji). W związku z tym mogą powstać problemy z określeniem, które pary powinny faktycznie podlegać redukcji. Poza tym w przypadku identyfikacji takich procesów, jak np. obiekty elektroakustyczne, dla których w charakterystyce amplitudowej występuje wiele pików i głębokich dolin blisko siebie, por. rys. 4.1, zastosowanie podejścia zaproponowanego w [127] jest niemożliwe, gdyż odpowiadające im zera i bieguny będą bliskie sobie, a więc ich dzielenie kompletnie zniekształciłoby charakterystykę częstotliwościową modelu.

Z przedstawionych wyżej względów w dalszym ciągu przyjmuje się, że poszukiwania stopni modelu należy rozpocząć od dołu. Ponieważ na ogół minimalizowane kryterium wyboru struktury wykazuje silną nieliniowość względem stopni wielomianów, stąd poszukiwań nie powinno się kończyć z chwilą osiągnięcia pierwszego minimum, ale przetestować także kilka kolejnych struktur. Wprawdzie w [117] pokazano, że dla modeli *ARMA* wystarczy szukać do pierwszego minimum, ale dla struktur bardziej złożonych takie podejście może doprowadzić do zbyt wczesnego zakończenia procedury poszukiwania.

Osobne zagadnienie stanowi problem oszacowania czasów opóźnienia d występujących w modelach torów sterowania. Wyznaczenie czasu opóźnienia na drodze analizy funkcji

korelacji wzajemnej wejścia i wyjścia obiektu, tak jak to zaproponowano w [45], może być ryzykowne w przypadku istnienia sprzężenia zwrotnego. Z kolei metoda wyznaczenia czasu opóźnienia na drodze analizy odpowiedzi impulsowej zaproponowana w [73] wymaga znajomości rzędu modelu. Można próbować oszacować czas opóźnienia na drodze analizy wartości współczynników wielomianu $B(z^{-1})$ uzyskanych dla modelu z czasem opóźnienia d_{min} i określenia, które z pierwszych współczynników tego wielomianu można uznać za nieistotne. Jednak procedura taka, aczkolwiek dość skuteczna, oparta jest na subiektywnej ocenie pomijalności współczynników wielomianu w modelu. Przeprowadzone badania symulacyjne wykazały, że najskuteczniejsza jest metoda najprostsza, polegająca na identyfikacji modeli *ARX* dla kolejnych czasów opóźnienia spośród zadanego zakresu $d_{min} \div d_{max}$ oraz najniższych założonych stopni wielomianów n_{min}, a następnie na wyborze takiego czasu opóźnienia $\hat{d}$, dla którego funkcja strat V jest najmniejsza. Metoda ta może wykazywać pewną tendencję do niewielkiego przeszacowania czasu opóźnienia, szczególnie wtedy, gdy kilka kolejnych współczynników wielomianu $B(z^{-1})$ w obiekcie jest małych w porównaniu z następnymi. Stąd przyjmuje się, że w następnych krokach procedury poszukiwania powinno się zaczynać od próby zmniejszania czasu opóźnienia w modelu.

Przedstawiony w dalszej części algorytm poszukiwania optymalnej (w sensie określonego kryterium) struktury modelu liniowego oparty został na zestawie reguł heurystycznych, wynikających głównie z przeprowadzonych licznych badań symulacyjnych. Podstawową zaletą tego typu badań jest to, że wiadomo, jaka jest struktura symulowanego procesu i stąd łatwo można ocenić wynik identyfikacji struktury.

Z przeprowadzonych badań nad algorytmem poszukiwania struktury dla obiektów *MISO* wynika podstawowy wniosek, że zarówno długość trwania procedury, jak i końcowy wynik zależą silnie od przyjętego punktu startowego. Ponieważ poszukiwanie struktury modelu *SISO* jest znacznie łatwiejsze ze względu na mniejszą liczbę zmiennych decyzyjnych, więc najlepiej, jeżeli tym punktem startowym jest struktura modelu *MISO* powstała ze złożenia struktur optymalnych uzyskanych dla poszczególnych torów sterowania traktowanych jako modele *SISO*, por. [58]. Tak więc cała procedura składa się z dwóch podstawowych etapów. Najpierw spośród zbioru danych $(\mathbf{U}_N, \mathbf{Y}_N)$ przetwarza się kolejno każdą parę danych $(\mathbf{U}_{k,N}, \mathbf{Y}_N)$ z osobna, gdzie $\mathbf{U}_{k,N}$ $(k = 1, 2, \ldots, q)$ oznacza ciąg danych dla k-tego wejścia. Umożliwia to uzyskanie q struktur modeli *SISO*. Następnie poszukuje się struktury modelu *MISO* opisującego zależność wyjścia od wszystkich pomiarowo dostępnych wejść, przy czym jako strukturę początkową przyjmuje się czasy opóźnienia i stopnie wielomianów $B(z^{-1})$ oraz $F(z^{-1})$ uzyskane w pierwszym kroku. Dla wielomianów wspólnych, tj. $A(z^{-1})$, $C(z^{-1})$ i $D(z^{-1})$, jako wartość startową przyjmuje się maksymalny stopień danego wielomianu uzyskany dla q modeli *SISO*.

Otrzymane w ten sposób stopnie wielomianów są na ogół zbyt duże (tzw. nadparametryzacja), co wynika stąd, że modele *SISO* usiłują opisać zależność wyjścia obiektu *MISO* od wszystkich jego wejść za pomocą modelowania zależności wejściowo-wyjściowej tylko dla jednej pary sygnałów. Stąd dalsze poszukiwania minimum przyjętego kryterium wyboru Q dla modelu *MISO* polegają głównie na zmniejszaniu stopni kolejnych wielomianów i testowaniu otrzymanych w ten sposób modeli. Przy okazji testowane są także sąsiednie wartości czasów opóźnienia. W efekcie, w stosunkowo niewielkiej liczbie kroków możliwe jest otrzymanie modelu *MISO* o optymalnej (w sensie testowanego kryterium) strukturze. Ponieważ przedstawiony algorytm jest oparty na regułach heurystycznych, więc nie ma pewności, że tak otrzymana struktura odpowiada minimum globalnemu. Przeprowadzone badania pokazały jednak, że na ogół osiągano minimum globalne, a w sytuacjach, gdy procedura zakończyła się na minimum lokalnym, to było ono bardzo blisko globalnego, przy czym różnice własności dynamicznych modeli dla obu struktur były pomijalnie małe, por. przykłady w [63].

Poszukiwanie struktury modelu *SISO* realizowane jest także za pomocą algorytmu opartego na regułach wynikających z doświadczeń symulacyjnych, które pokazują, że na ogół optymalne stopnie poszczególnych wielomianów nie różnią się znacznie między sobą. Stąd podstawową cechą algorytmu jest rozpoczęcie poszukiwań od przyjęcia takiego samego stopnia n dla wszystkich wielomianów w modelu. Następnie, po znalezieniu takiego $\hat{n}$, dla którego osiągnięto minimalną wartość kryterium, następuje testowanie modeli o stopniach wielomianów i czasach opóźnienia wokół poprzednio uzyskanych wartości. Kolejność testowanych wielomianów nie jest przypadkowa, ale wynika z przesłanek opartych na analizie wpływu tych wielomianów na własności modelu oraz z doświadczenia.

Korzystając z konwencji strukturalnego języka programowania (symbol ← oznacza tu wyznaczenie wartości lub instrukcję przypisania) przedstawiono poniżej bardziej szczegółowy zapis algorytmów poszukiwania struktur modeli liniowych *SISO* i *MISO*. Algorytmy przedstawiono dla struktury ogólnej, w przypadku typów szczególnych, jak np. *ARX*, *ARMAX*, itd., por. 4.2.1, odpowiednie stopnie wielomianów są z definicji równe zero. Warunkami początkowymi dla procedury są zadeklarowane przez użytkownika: typ modelu $\mathcal{F}$, zakresy poszukiwań dla czasów opóźnienia $d_{\min}, d_{\max}$ oraz stopni wielomianów $n_{\min}, n_{\max}$, a także kryterium wyboru Q (*AIC*, *BIC* lub inne, por. 5.3.1). Element struktury wybrany zgodnie z minimum kryterium Q oznaczono daszkiem. W celu zwiększenia czytelności w nawiasach podano tylko te elementy wektora μ, których wpływ na wartość Q w danym kroku procedury jest testowany.


```
BEGIN (* Procedura poszukiwania struktury modelu SISO *)
  (* Oszacowanie czasu opóźnienia d *)
  FOR d := d_min TO d_max
    V(d) ← ARX(d, dA = dB = n_min) (* Wyznacz model ARX i wartość funkcji strat *)
  END_FOR
  d̂ ← arg min V(d)
  (* Wstępne oszacowanie stopnia modelu *)
  FOR n := n_min TO n_max (* por. Uwaga 3 *)
    Q(n) ← M(d̂, dA = dB = dC = dD = dF = n) (* Wyznacz model i wartość kryterium Q*)
  END_FOR
  n̂ ← arg min Q(n)
  dÂ = dB̂ = dĈ = dD̂ = dF̂ = n̂
  (* Poszukiwanie struktury μ na drodze iteracyjnego testowania wartości Q dla kolejno
     zmienianych d, dD, dC, dB, dF, dA *)
  REPEAT
    IF NOT (zwiększono d) AND (d̂ > 1) THEN
      Q(d̂ − 1) ← M(d̂ − 1) (* Wyznacz model i wartość Q dla mniejszego d *)
      IF (Q(d̂) > Q(d̂ − 1)) THEN
        d̂ ← d̂ − 1, zmniejszono d (* Zmniejsz czas opóźnienia *)
      ELSE (* por. Uwaga 1 *)
        (* Wyznacz model i wartość Q dla mniejszego d i większego dB *)
        Q(d̂ − 1, dB̂ + 1) ← M(d̂ − 1, dB̂ + 1)
        IF (Q(d̂, dB̂) > Q(d̂ − 1, dB̂ + 1)) THEN
          (* Zmniejsz czas opóźnienia i zwiększ stopień wielomianu B *)
          d̂ ← d̂ − 1, dB̂ ← dB̂ + 1, zmniejszono d, zwiększono dB
        END_IF
      END_IF
    END_IF
    IF NOT (zmniejszono d) THEN
      Q(d̂ + 1) ← M(d̂ + 1) (* Wyznacz model i wartość Q dla większego d *)
      IF (Q(d̂) > Q(d̂ + 1)) THEN
        d̂ ← d̂ + 1, zwiększono d (* Zwiększ czas opóźnienia *)
      ELSE (* por. Uwaga 1 *)
```

(* Wyznacz model i wartość Q dlawiększego d i mniejszego dB *)

$Q(\hat{d}+1,\ d\hat{B}-1) \leftarrow M(\hat{d}+1,\ d\hat{B}-1)$

IF $(Q(\hat{d},\ d\hat{B}) > Q(\hat{d}+1,\ d\hat{B}-1))$ THEN

(* Zwiększ czas opóźnienia i zmniejsz stopień wielomianu B *)

$\hat{d} \leftarrow \hat{d}+1,\ d\hat{B} \leftarrow d\hat{B}-1$, *zwiększono d*

END_IF

END_IF

END_IF

(* Badaj kolejne wielomiany *)

IF $(d\hat{D} > 0)$ THEN

$Q(d\hat{D}-1) \leftarrow M(d\hat{D}-1)$ (* Wyznacz model i wartość Q dla mniejszego dD *)

IF $(Q(d\hat{D}) > Q(d\hat{D}-1))$ THEN

$d\hat{D} \leftarrow d\hat{D}-1$ (* Zmniejsz stopień wielomianu D *)

END_IF

END_IF

IF $(d\hat{C} > 0)$ THEN

$Q(d\hat{C}-1) \leftarrow M(d\hat{C}-1)$ (* Wyznacz model i wartość Q dla mniejszego dC *)

IF $(Q(d\hat{C}) > Q(d\hat{C}-1))$ THEN

$d\hat{C} \leftarrow d\hat{C}-1$ (* Zmniejsz stopień wielomianu C *)

END_IF

END_IF

IF NOT (*zwiększono dB*) AND $(d\hat{B} > 0)$ THEN (* por. Uwaga 1 *)

$Q(d\hat{B}-1) \leftarrow M(d\hat{B}-1)$ (* Wyznacz model i wartość Q dla mniejszego dB *)

IF $(Q(d\hat{B}) > Q(d\hat{B}-1))$ THEN

$d\hat{B} \leftarrow d\hat{B}-1$ (* Zmniejsz stopień wielomianu B *)

END_IF

END_IF

IF $(d\hat{F} > 0)$ THEN

$Q(d\hat{F}-1) \leftarrow M(d\hat{F}-1)$ (* Wyznacz model i wartość Q dla mniejszego dF *)

IF $(Q(d\hat{F}) > Q(d\hat{F}-1))$ THEN

$d\hat{F} \leftarrow d\hat{F}-1$ (* Zmniejsz stopień wielomianu F *)

END_IF

END_IF

IF $(d\hat{A} > 0)$ THEN

$Q(d\hat{A}-1) \leftarrow M(d\hat{A}-1)$ (* Wyznacz model i wartość Q dla mniejszego dA *)

IF $(Q(d\hat{A}) > Q(d\hat{A}-1))$ THEN

$d\hat{A} \leftarrow d\hat{A}-1$, *zmniejszono dA* (* Zmniejsz stopień wielomianu A *)

END_IF

END_IF

UNTIL NOT (*zmieniono strukturę*)

IF $(d\hat{A} < n_{\max})$ AND NOT (*zmniejszono dA*) THEN (* por. Uwaga 2 *)

$Q(d\hat{A}+1) \leftarrow M(d\hat{A}+1)$ (* Wyznacz model i wartość Q dla większego dA *)

IF $(Q(d\hat{A}) > Q(d\hat{A}+1))$ THEN

$d\hat{A} \leftarrow d\hat{A}+1$ (* Zwiększ stopień wielomianu A *)

END_IF

END_IF

(* Przyjęcie struktury optymalnej *)

$\mu \leftarrow (\hat{d},\ d\hat{A},\ d\hat{B},\ d\hat{F},\ d\hat{C},\ d\hat{D})$

END (* Koniec procedury poszukiwania struktury modelu SISO *)

BEGIN (* Procedura poszukiwania struktury modelu MISO *)

(* Określenie punktu startowego *)

FOR $k := 1$ TO q (* Wyznacz optymalną strukturę k-tego toru sterowania *)

$(\hat{d}_k, d\hat{B}_k, d\hat{F}_k) \leftarrow SISO(\mu_k)$

END_FOR

(* Dla wielomianów wspólnych wybierz stopnie największe *)

$d\hat{A} \leftarrow \max_k d\hat{A}_k,\ d\hat{C} \leftarrow \max_k d\hat{C}_k,\ d\hat{D} \leftarrow \max_k d\hat{D}_k$

(* Poszukiwanie struktury μ na drodze iteracyjnego testowania wartości Q dla kolejno zmienianych stopni wielomianów i czasów opóźnienia *)

REPEAT

IF $(d\hat{A} > 0)$ THEN

$Q(d\hat{A}-1) \leftarrow M(d\hat{A}-1)$ (* Wyznacz model i wartość Q dla mniejszego dA *)

IF $(Q(d\hat{A}) > Q(d\hat{A}-1))$ THEN

$d\hat{A} \leftarrow d\hat{A}-1$, (* Zmniejsz stopień wielomianu A *)

END_IF

END_IF

IF ($d\hat{D} > 0$) THEN

$Q(d\hat{D}-1) \leftarrow M(d\hat{D}-1)$ (* Wyznacz model i wartość Q dla mniejszego dD *)

IF ($Q(d\hat{D}) > Q(d\hat{D}-1)$) THEN

$d\hat{D} \leftarrow d\hat{D}-1$, (* Zmniejsz stopień wielomianu D *)

END_IF

END_IF

IF ($d\hat{C} > 0$) THEN

$Q(d\hat{C}-1) \leftarrow M(d\hat{C}-1)$ (* Wyznacz model i wartość Q dla mniejszego dC *)

IF ($Q(d\hat{C}) > Q(d\hat{C}-1)$) THEN

$d\hat{C} \leftarrow d\hat{C}-1$, (* Zmniejsz stopień wielomianu C *)

END_IF

END_IF

(* Kolejne tory sterowania *)

FOR $k := 1$ TO q

IF ($d\hat{F}_k > 0$) THEN

$Q(d\hat{F}_k-1) \leftarrow M(d\hat{F}_k-1)$ (* Wyznacz model i wartość Q dla mniejszego dF_k *)

IF ($Q(d\hat{F}_k) > Q(d\hat{F}_k-1)$) THEN

$d\hat{F}_k \leftarrow d\hat{F}_k-1$, *zmniejszono* dF_k (* Zmniejsz stopień wielomianu F_k *)

END_IF

END_IF

IF NOT (*zwiększono dk*) AND ($\hat{\partial}_k > 1$) THEN

$Q(\hat{\partial}_k-1) \leftarrow M(\hat{\partial}_k-1)$ (* Wyznacz model i wartość Q dla mniejszego d_k *)

IF ($Q(\hat{\partial}_k) > Q(\hat{\partial}_k-1)$) THEN

$\hat{\partial}_k \leftarrow \hat{\partial}_k-1$, *zmniejszono* d_k (* zmniejsz czas opóźnienia *)

ELSE (* Uwaga 1 *)

(* Wyznacz model i wartość Q dla mniejszego d_k i większego dB_k *)

$Q(\hat{\partial}_k-1, d\hat{B}_k+1) \leftarrow M(\hat{\partial}_k-1, d\hat{B}_k+1)$

IF ($Q(\hat{\partial}_k, d\hat{B}_k) > Q(\hat{\partial}_k-1, d\hat{B}_k+1)$) THEN

(* Zmniejsz czas opóźnienia i zwiększ stopień wielomianu B *)

$\hat{\partial}_k \leftarrow \hat{\partial}_k-1$, $d\hat{B}_k \leftarrow d\hat{B}_k+1$, *zmniejszono* d_k, *zwiększono* dB_k

END_IF

END_IF

END_IF

IF NOT (*zmniejszono* d_k) THEN

$Q(\hat{\partial}_k+1) \leftarrow M(\hat{\partial}_k+1)$ (* Wyznacz model i wartość Q dla większego d_k *)

IF ($Q(\hat{\partial}_k) > Q(\hat{\partial}_k+1)$) THEN

$\hat{\partial}_k \leftarrow \hat{\partial}_k+1$, *zwiększono* d_k (* Zwiększ czas opóźnienia *)

ELSE (* por. Uwaga 1 *)

(* Wyznacz model i wartość Q dla większego d_k i mniejszego dB_k *)

$Q(\hat{\partial}_k+1, d\hat{B}_k-1) \leftarrow M(\hat{\partial}_k+1, d\hat{B}_k-1)$

IF ($Q(\hat{\partial}_k, d\hat{B}_k) > Q(\hat{\partial}_k+1, d\hat{B}_k-1)$) THEN

(* Zwiększ czas opóźnienia i zmniejsz stopień wielomianu B_k *)

$\hat{\partial}_k \leftarrow \hat{\partial}_k+1$, $d\hat{B}_k \leftarrow d\hat{B}_k-1$, *zwiększono* d_k

END_IF

END_IF

END_IF

IF NOT (*zwiększono dBk*) AND ($d\hat{B}_k > 0$) THEN (* por. Uwaga 1 *)

$Q(d\hat{B}_k-1) \leftarrow M(d\hat{B}_k-1)$ (* Wyznacz model i wartość Q dla mniejszego dB_k *)

IF ($Q(d\hat{B}_k) > Q(d\hat{B}_k-1)$) THEN

$d\hat{B}_k \leftarrow d\hat{B}_k-1$, (* Zmniejsz stopień wielomianu B_k *)

END_IF

END_IF

IF NOT (*zmniejszono dFk*) THEN

$Q(d\hat{F}_k+1) \leftarrow M(d\hat{F}_k+1)$ (* Wyznacz model i wartość Q dla większego dF_k *)

IF ($Q(d\hat{F}_k) > Q(d\hat{F}_k+1)$) THEN

$d\hat{F}_k \leftarrow d\hat{F}_k+1$, (* Zwiększ stopień wielomianu F_k *)

END_IF

END_IF

END_FOR

UNTIL NOT (*zmieniono strukturę*)

$\hat{\partial} = [\hat{\partial}_1 \quad \hat{\partial}_2 \quad \ldots \quad \hat{\partial}_m] \quad d\hat{B} = [d\hat{B}_1 \quad d\hat{B}_2 \quad \ldots \quad d\hat{B}_m] \quad d\hat{F} = [d\hat{F}_1 \quad d\hat{F}_2 \quad \ldots \quad d\hat{F}_m]$

(* Przyjęcie struktury optymalnej *)

$p \leftarrow (\hat{\partial}, d\hat{A}, d\hat{B}, d\hat{F}, d\hat{C}, d\hat{D})$

END (* Koniec procedury poszukiwania struktury modelu MISO *)

Poniżej przedstawiono dodatkowe uwagi dotyczące powyższych algorytmów.

Uwaga 1. Zmniejszenie czasu opóźnienia w modelu przy niezmienionym stopniu *dB* powoduje, że z równania różnicowego znika składnik $u(i-d-dB)$, natomiast pojawia się składnik $u(i-d+1)$. W związku z tym, jeżeli model dla zmniejszonego *d* nie okaże się lepszy, to wtedy testuje się model zawierający oba składniki, tj. dla zmniejszonego *d* oraz zwiększonego *dB*. Podobnie, zwiększenie *d* przy niezmienionym *dB* powoduje, że z równania różnicowego znika składnik $u(i-d)$, natomiast pojawia się składnik $u(i-d-dB-1)$. W związku z tym, jeżeli model dla zwiększonego *d* nie okaże się lepszy, to wtedy testuje się model pozbawiony obu składników, tj. dla zwiększonego *d* oraz zmniejszonego *dB*.

Uwaga 2. Ponieważ na ogół wielomian autoregresji $A(z^{-1})$ ma największy wpływ na własności dynamiczne modelu SISO, więc na zakończenie procedury testuje się jeszcze jeden model dla zwiększonego stopnia *dA*, pod warunkiem oczywiście, że w trakcie poprzedniej iteracji zmniejszania stopni wielomianów *dA* nie zostało zmniejszone.

Uwaga 3. W przedstawionych wyżej procedurach automatycznego poszukiwania struktury modelu przyjęto w domyśle, że minimalizacja kryterium Q jest odpowiednikiem uzyskania modelu *odpowiedniego do celu*. Jednak po uzyskaniu modelu powinien on zostać poddany odpowiednim testom, które umożliwią jego walidację i falsyfikację, co jest tematem następnego punktu. Jeśli natomiast możliwe byłoby określenie kryterium odpowiedniości dla celu w postaci $\Delta(\hat{M} \mid \mathrm{M}) \le \delta$, por. 5.1, to wtedy należałoby wybrać najprostszy model spełniający to kryterium. Zgodnie więc z ogólną procedurą przedstawioną w 5.2.1, pętla FOR powinna zostać zastąpiona pętlą REPEAT:

$n := n_{\min} - 1$
REPEAT
  $n \leftarrow n + 1$
  $M(\hat{\partial}, dA = dB = dC = dD = dF = n)$ (* Wyznacz model *)
  IF $\Delta(\hat{M} \mid \mathbf{M}) \le \delta$ THEN *model odpowiedni do celu*
  ELSE *model nieodpowiedni do celu*
  END_IF
UNTIL (*model nieodpowiedni do celu*) AND ($n < n_{\max}$)
IF *model odpowiedni do celu* THEN
  $d\hat{A} = d\hat{B} = d\hat{C} = d\hat{D} = d\hat{F} = n$
ELSE *nie można wyznaczyć modelu odpowiedniego do celu*
END_IF

Negatywny wynik poszukiwań wymaga wykonania zewnętrznej pętli w ogólnej procedurze identyfikacji, tzn. przyjęcia innej formy modelu $\mathcal{F}$ lub zadeklarowania większego $n_{\max}$, z czym może się wiązać także konieczność wykonania nowego eksperymentu $R_{\max}$, por. 5.2.1.

Przedstawione wyżej algorytmy poszukiwania struktury liniowego modelu parametrycznego były szeroko testowane, zarówno na danych symulowanych, jak i pochodzących z rzeczywistych obiektów. Wykazały one swą użyteczność i wysoką skuteczność dla identyfikacji bardzo różnych procesów, aczkolwiek opracowane zostały w postaci reguł heurystycznych, bazujących na doświadczeniu własnym autora. W dopracowaniu algorytmów dużą rolę odgrywała możliwość śledzenia pracy systemu *MULTI-EDIP* dzięki prezentacji w oknie raportu informacji o kolejnych testowanych strukturach (por. dodatek C). W następnym punkcie pokazano przykład działania algorytmu dla danych symulowanych, natomiast w rozdziale 7 przedstawiono wyniki działania algorytmu w identyfikacji przykładowego obiektu elektroakustycznego.

5.3.3. Przykład symulacyjny

Działanie algorytmu automatycznego poszukiwania struktury modelu zostanie przedstawione na przykładzie symulacji obiektu o dwóch wejściach, zakłócanego losowo. Symulowany był obiekt typu *BJ* opisany równaniem:

$$y(i) = z^{-3}\frac{0.15+0.1z^{-1}+0.1z^{-2}}{1-1.3z^{-1}+0.65z^{-2}}u_1(i) + z^{-1}\frac{0.1+0.2z^{-1}}{1-1.15z^{-1}+0.5z^{-2}}u_2(i) + \frac{1}{1-1.5z^{-1}+0.7z^{-2}}e(i),$$

pobudzany *wektorowymi sygnałami wielosinusoidalnymi ortogonalnymi* (WSWO), por. [28], dla liczby sinusów 256 i losowej fazy. Amplitudy sinusów dobrano tak, aby odchylenia standardowe (OS) obu sygnałów wejściowych wynosiły 1.0, co daje OS cząstkowego wyjścia pierwszego toru sterowania ok. 0.65, a drugiego ok. 0.5. Natomiast OS białego szumu $e(i)$ wynosiło 0.167, co daje OS zakłócenia sprowadzonego na wyjście ok. 0.5. Warto zwrócić uwagę na to, że drugi tor sterowania jest nieminimalnofazowy, co może utrudniać wyznaczenie czasu opóźnienia w tym torze. Poza tym, własności dynamiczne w poszczególnych torach są zbliżone (są to elementy oscylacyjne drugiego stopnia), różnią się jednak częstotliwościami własnymi i współczynnikami tłumienia, co także nie ułatwia identyfikacji.

Badania poszukiwania struktury modelu *BJ* ($[d_1\ d_2]$, $[dB_1\ dB_2]$, $[dF_1\ dF_2]$, dC, dD) przeprowadzono dla dziesięciu serii pomiarowych liczących $N = 512$ danych każda. Zakres poszukiwań dla czasów opóźnienia ustalono na $d_{min} = 1$, $d_{max} = 5$ oraz podobnie dla stopni wielomianów $n_{min} = 1$ i $n_{max} = 5$. Modele identyfikowane były z wykorzystaniem rekurencyjnej metody błędu predykcji *RPE*, por. dodatek A.2. Poniżej omówiono otrzymane wyniki.

1. Kryterium *BIC*

Dziewięciokrotnie uzyskano wynik w postaci struktury modelu *BJ* $\underline{\alpha}$ = ([3 1], [2 1], [2 2], 0, 2), a więc dokładnie takiej, jaka była symulowana. W jednym przypadku uzyskano model o strukturze $\underline{\alpha}$ = ([3 1], [2 1], [2 2], 2, 2), a więc o $d\hat{C} = 2$ zamiast 0, dla której wartość *BIC* = -1734.8. Dla struktury poprawnej *BIC* = -1741.5, jednak dla $dC = 1$ wartość *BIC* = -1734.2, a więc zgodnie z zaproponowanym algorytmem, system *MULTI-EDIP* zakończył poszukiwania dla $d\hat{C} = 2$. Przykład ten pokazuje, że ze względu na nieliniowe własności kryterium wyboru poszukiwanie struktury może się zatrzymać na minimum lokalnym. Jednak przypadki takie nie zdarzają się często, poza tym wybrana wtedy struktura modelu na ogół różni się nieznacznie od struktury optymalnej (w sensie tego kryterium).

Warto tu podkreślić, że zaproponowany sposób oszacowania czasu opóźnienia za każdym razem pozwalał na uzyskanie poprawnych wartości opóźnień $\hat{d}$ w każdym z torów sterowania, mimo nieminimalnofazowości występującej w drugim torze. Podobnie poszukiwania struktur (*d*, *dB*, *dF*, *dC*, *dD*) modeli *SISO* dla par (u_1, y) oraz (u_2, y) dawały odpowiednio (3, 2, 2, 1, 2) dla toru pierwszego oraz (1, 1, 2, 2, 2) dla drugiego, a więc bardzo bliskie poprawnym. Jest to o tyle istotne, że rozpoczęcie poszukiwań struktury modelu *MISO* od wartości zbliżonej do prawidłowej - tu było to ([3 1], [2 1], [2 2], 2, 2) – pozwala w stosunkowo niewielkiej liczbie kroków znaleźć strukturę optymalną. W sumie w prezentowanym przykładzie procedura wymagała identyfikacji około 20 modeli *SISO* dla pierwszego toru, 15 modeli *SISO* dla drugiego oraz ponad 30 modeli *MISO*. W przypadku przeglądu zupełnego trzeba by przetestować około 400000 modeli *MISO* dla przyjętych granic poszukiwania.

2. Kryterium *AIC*

Poszukiwania punktu startowego na drodze identyfikacji modeli *SISO* dały dokładnie takie same wyniki jak dla kryterium *BIC*. Jednak w żadnym z dziesięciu przypadków optymalizacja *AIC* nie doprowadziła do uzyskania struktury modelu identycznej z symulowaną. W każdym eksperymencie otrzymano model o liczbie parametrów większej, niż była konieczna, przy czym w każdym przypadku nie był to efekt błędnego zakończenia procedury, ale dla takich właśnie struktur występowała minimalna wartość *AIC*. Na przykład, dla jednej z serii danych uzyskano model o strukturze $\underline{\alpha}$ = ([1 1], [4 3], [2 2], 2, 2), dla którego *AIC* = -1819.1, podczas gdy dla struktury równej symulowanej wartość *AIC* była większa i wynosiła -1815.6. Nadmiarowe parametry (od jednego do trzech) występowały dla któregoś z wielomianów $B(z^{-1})$ lub $C(z^{-1})$. W omawianym przypadku na drodze analizy

wartości parametrów wielomianu $B_1(z^{-1})$ można by było dokonać próby korekty struktury modelu poprzez zwiększenie czasu opóźnienia w pierwszym torze o 2, albowiem zidentyfikowane parametry $\hat{b}_{1,0}$ oraz $\hat{b}_{1,1}$, występujące przy wartościach u_1 przesuniętych wstecz o mniej niż wynosił symulowany czas opóźnienia, były znacznie mniejsze niż następne. Jednak tego typu analiza byłaby bardzo trudna do przeprowadzenia w sposób automatyczny, gdyż trudno by było sformułować uniwersalne kryterium pozwalające na tego typu redukcję modelu.

3. Kryterium wariancji błędów predykcji dla danych do testowania

Poszukiwanie struktury modelu minimalizującej błąd predykcji, obliczony dla zbioru danych innego, niż wzięty do estymacji, także dał za każdym razem wynik w postaci modelu zawierającego od jednego do kilku parametrów więcej, niż było symulowanych. Parametry te występowały dla któregoś z wielomianów $B(z^{-1})$ lub $C(z^{-1})$, podobnie jak dla kryterium *AIC*.

4. Kryterium *FOE*

Kryterium oparte na minimalizacji błędu wyjścia (5.13) doprowadziło za każdym razem do uzyskania modelu o strukturze *OE*, a więc posiadającego $d\hat{C} = d\hat{D} = 0$. Można było tego w pewien sposób oczekiwać, jako że istotą tego kryterium jest minimalizacja błędu wyjścia, a nie predykcji. Jednocześnie na ogół struktura modeli torów sterowania była prawidłowa, aczkolwiek wystąpiły przypadki, gdy uzyskano model zawierający dodatkowe zero i biegun, np. $d\hat{B}_1 = d\hat{F}_1 = 3$ zamiast 2.

W przypadku kryterium *FOE* nieco lepsze wyniki osiągnięto stosując do estymacji parametrów modelu zamiast metody *RPE* metodę zaproponowaną w [24], polegającą na iteracyjnej identyfikacji modeli torów sterowania za pomocą metody zmiennych pomocniczych (instrumentalnych). Dla tak otrzymanych modeli torów sterowania obliczano następnie błąd wyjścia, dla którego z kolei dokonywano identyfikacji modelu ciągu czasowego *ARMA*, przy czym do wyboru struktury tego modelu stosowano już kryterium *BIC*. Taka dwustopniowa metoda identyfikacji modelu *BJ* umożliwiała otrzymanie modeli torów sterowania o podobnej dokładności jak dla *RPE* z kryterium *FOE*, natomiast uzyskane modele toru zakłócenia były albo typu *AR* o $d\hat{D} = 2$, albo *ARMA* o $d\hat{C} = d\hat{D} = 2$.

Oczywiście, trudno jest wyciągać ogólne wnioski na podstawie pojedynczych badań symulacyjnych. Jednak wieloletnie doświadczenia uzyskane w trakcie stosowania przedstawionego w npoprzednim punkcie algorytmu pokazały jego zasadność, przy czym

generalnie najlepsze wyniki osiągano właśnie dla kryterium *BIC*. Szczególnie dotyczy to problemów identyfikacji modeli *MISO*.

5.4. Wspomaganie w testowaniu i analizie modeli liniowych

Stosując zasadę wyboru struktury modelu na drodze minimalizacji przyjętego kryterium Q zakłada się milcząco, że w ten sposób można uzyskać model *odpowiedni do celu*. Jednak po pierwsze, nie ma pewności, że kryterium takie rzeczywiście jest odpowiednie do celu (np. w identyfikacji modelu obiektu elektroakustycznego dla celów aktywnego tłumienia hałasu, por. 5.1.4), a po drugie, nie ma pewności, że w zadeklarowanym zakresie poszukiwań $n_{\max}$ dla wybranej formy $\mathcal{F}$ znajduje się model odpowiedni do celu. Stąd niezwykle ważne jest, aby po zakończeniu procedury poszukiwań uzyskany model poddać odpowiednim testom, w wyniku których model zostanie zaakceptowany lub odrzucony.

Bezsprzecznie o odpowiedniości modelu do celu decyduje test *zastosowania* modelu do zadania, dla którego był przeznaczony. Niemniej jednak przed przystąpieniem do testu zastosowań dobrze jest poddać model pewnym testom formalnym, które umożliwią odrzucenie modeli ewidentnie złych lub spośród kilku modeli pozwolą wybrać model najlepszy w określonym sensie.

Istotny jest tu dobór odpowiednich testów. W punkcie 5.1 omówiono podstawowe zasady dotyczące testowania modelu stochastycznego. W literaturze przedmiotu można znaleźć wiele mniej lub bardziej wyszukanych metod testowania modelu. Metody te w głównej mierze opierają się na teorii testowania hipotez i służą przede wszystkim falsyfikacji modelu. Pakiet programowy do wspomagania identyfikacji powinien zawierać wiele opcji w tym zakresie, jako że im więcej testów, tym bardziej zawęża się zbiór modeli niesfalsyfikowanych $\mathcal{N}$. W szczególności dotyczy to testów białości błędów predykcji czy testów porównawczych dla różnych struktur, które wykorzystywane są w procedurze identyfikacji struktury modelu.

Jednak oprócz testów formalnych, oprogramowanie powinno dostarczać także pewnej liczby testów mniej formalnych, ale za to bardziej poglądowych. Trzeba się bowiem liczyć z tym, że użytkownik takiego oprogramowania nie musi być specjalistą w zakresie statystyki i oczekuje przede wszystkim takich testów, które umożliwiają ocenę modelu w sposób intuicyjnie zrozumiały. Jednocześnie należy mu uświadomić fakt, że model nie dlatego jest „dobry", że jest „najlepszy" w wybranej klasie modeli, ale dlatego, że przeszedł odpowiednie testy. Powinien więc model oceniać pod różnymi kątami, aby móc określić, że może być on wystarczająco dobry dla postawionego celu.

W dalszej części przedstawiono podstawowy zestaw testów wizualizacyjnych, które zaproponowano w systemie *MULTI-EDIP* i których szczegółowe omówienie można znaleźć w [101]. Oprócz tego omówiono testy umożliwiające badanie uwarunkowania danych pomiarowych oraz testy pomocnicze pozwalające na określenie przyczyn złego uwarunkowania. Zagadnienie to jest rzadko prezentowane w literaturze na temat identyfikacji, a tymczasem ma ono podstawowe znaczenie dla uzyskania poprawnych wyników identyfikacji – szczególnie w przypadku obiektów o wielu wejściach. Niespełnienie odpowiednich warunków dobrego uwarunkowania może bowiem łatwo sprowadzić na manowce, szczególnie niedoświadczonego użytkownika oprogramowania do identyfikacji.

5.4.1. Testy wizualizacyjne

Istota tych testów polega na wyświetlaniu określonych charakterystyk modelu w celu ich wizualnej i zdroworozsądkowej oceny. Pomimo prostoty i subiektywnego charakteru testy te należą do kanonu metod testowania i analizy wyników identyfikacji, gdyż skutecznie umożliwiają wykrywanie modeli ewidentnie złych czy obarczonych błędami grubymi, bez konieczności sięgania do testów bardziej wyrafinowanych.

Do podstawowych testów wizualizacyjnych należą testy symulacji i predykcji, polegające na wizualnym porównaniu przebiegów czasowych zmierzonego ciągu wyjściowego obiektu $\mathbf{y}(i)$ z:

- *symulowanym* ciągiem wyjściowym testowanego modelu:

$$\hat{y}(i) = \sum_{k=1}^{q} \hat{G}_k\left(z^{-1}\right) u_k(i), \tag{5.14}$$

- *jednokrokową predykcją* ciągu wyjściowego, obliczoną na podstawie modelu i danych:

$$\hat{y}(i \mid i-1) = \hat{H}^{-1}\left(z^{-1}\right) \sum_{k=1}^{q} \hat{G}_k\left(z^{-1}\right) u_k(i) + \left[1 - \hat{H}^{-1}\left(z^{-1}\right)\right] y(i). \tag{5.15}$$

Oba testy są intuicyjne i pouczające oraz nie wymagają znajomości statystyki. Zaleca się, aby testy te wykonywane były na zbiorze *danych do testowania*, różnym od zbioru użytego do estymacji modelu. Dotyczy to szczególnie testu predykcji w sytuacji, gdy kryterium aproksymacji modelu oparte zostało na minimalizacji błędów predykcji, jak np. w metodzie *ML*. W takim przypadku szczególnie warto stosować test symulacji. Poza tym jeżeli dane zostały zebrane dla krótkiego okresu próbkowania, to wartość testu predykcji może być ograniczona, albowiem wtedy nawet doskonała predykcja jednokrokowa może niewiele mówić o jakości

modelu i wtedy należy raczej próbować oceniać model na podstawie predykcji obliczonej dla dłuższego horyzontu. Z oczywistych względów testu symulacji nie stosuje się w identyfikacji ciągów czasowych.

Kolejną grupę stanowią testy umożliwiające wizualną ocenę *własności dynamicznych* modelu, a w szczególności porównanie tych własności z modelami uzyskanymi metodami nieparametrycznymi lub z wiedzą *a priori* o własnościach identyfikowanego procesu. W przypadku obiektu ocenie podlegać mogą odpowiedzi impulsowe i skokowe torów sterowania, które można porównać np. z odpowiedziami zmierzonymi (o ile możliwe było przeprowadzenie tego typu eksperymentów) lub obliczonymi za pomocą metod nieparametrycznych, np. metodą korelacyjną [62]. Podobnie można porównać charakterystyki częstotliwościowe (wykresy Bode lub na płaszczyźnie Nyquista) modeli torów sterowania z charakterystykami uzyskanymi metodą periodogramową lub korelogramową (Blackmana-Tuckeya), por. np. [101].

W przypadku ciągów czasowych ocenie powinna podlegać funkcja gęstości widmowej mocy, którą można porównać z gęstością widmową oszacowaną za pomocą metody periodogramowej lub korelogramowej, w szczególności ze względu na liczbę pików rezonansowych czy dolin antyrezonansowych lub podobieństwo kształtów. Z kolei porównanie funkcji gęstości widmowej mocy obliczonych dla modeli parametrycznych o sąsiadujących strukturach może także dostarczyć pewnych wskazówek na temat nadparametryzacji w modelu.

Ponieważ wyznaczona ocena parametrów modelu procesu zakłócanego stochastycznie jest tylko jedną konkretną realizacją wielowymiarowej zmiennej losowej, stąd w przypadku silnych zakłóceń i krótkich ciągów danych inne realizacje mogą się znacznie różnić od wartości wyznaczonej. Miarą tej różnicy jest *krańcowa macierz kowariancji parametrów*, której analiza jednak nie jest prosta ze względu na jej niediagonalny charakter. Natomiast korzystając z technik Monte-Carlo można na podstawie tej macierzy wygenerować zbiór teoretycznie możliwych wektorów parametrów modelu, które z kolei służą do wykreślenia rodziny nakładających się charakterystyk odpowiadających temu zbiorowi, por. [101]. W ten sposób uzyskuje się wizualną możliwość oceny *teoretycznej wrażliwości modelu na losowość danych*. Metoda ta może być stosowana dla przedstawionych wyżej charakterystyk własności dynamicznych modelu, tj. odpowiedzi impulsowych i skokowych oraz charakterystyk częstotliwościowych. Duża wrażliwość na losowość danych może być także spowodowana słabym uwarunkowaniem równań estymatora, stąd istotne jest również badanie uwarunkowania numerycznego zadania identyfikacji.

5.4.2. Badanie uwarunkowania numerycznego

W analizie wyników identyfikacji bardzo istotne jest sprawdzenie uwarunkowania równań estymatora, ponieważ złe uwarunkowanie świadczy o tym, że być może struktura, a na pewno parametry modelu nie są poprawne. W takiej sytuacji bowiem występuje duża wrażliwość wyników identyfikacji na losowość danych pomiarowych. Wobec tego należy znaleźć i usunąć przyczyny złego uwarunkowania.

Test uwarunkowania równań estymatora można przeprowadzić albo badając stosunek wartości własnych: maksymalnej do minimalnej, obliczonych dla końcowej macierzy danych, albo badając wykres zmienności minimalnej wartości własnej macierzy w zależności od liczby pomiarów N, por. [101]. Badanie minimalnej wartości własnej może być także wykorzystane jako test pomocniczy w identyfikacji struktury modeli *MIMO*, por. [103].

W systemie *MULTI-EDIP* wykonywane są oba testy, przy czym w celu automatycznego stwierdzenia, czy układ równań estymatora jest dobrze uwarunkowany, przyjęto arbitralnie wartość 10^8 jako granicę dla stosunku wartości własnych. W przypadku testu przebiegu minimalnej wartości własnej dokonuje się jego aproksymacji trendem wielomianowym (3.1) pierwszego stopnia i bada zmienność współczynnika c_1 tego trendu.

W przypadku stwierdzenia, że uwarunkowanie macierzy danych pomiarowych jest złe, można zastosować szereg testów pomocniczych, które pozwolą wykryć przyczynę słabego uwarunkowania. Może ono wynikać z nieprawidłowo przeprowadzonego eksperymentu identyfikacyjnego lub ze złej struktury modelu. Do podstawowych testów pomocniczych należą testy badające stopień uwarunkowania (definiowany jako stosunek wartości osobliwych albo własnych) odpowiednich macierzy korelacji.

I tak, badanie stopnia uwarunkowania *macierzy autokorelacji wyjść* ρ_{yy}, por. dodatek B, o wymiarach $dA + dF_{max}$, gdzie dF_{max} jest maksymalnym stopniem wielomianów $F_k(z^{-1})$, pozwala na wykrycie zbyt krótkiego okresu próbkowania (zwłaszcza gdy wartość funkcji autokorelacji dla przesunięcia 1 jest bliska wartości dla przesunięcia 0), istnienia *offsetu* w danych pomiarowych (gdy odchylenie standardowe wyjścia jest porównywalne z jego wartością średnią) lub zbyt wysokiego stopnia obu wielomianów autoregresji.

Natomiast badanie uwarunkowania odpowiednich macierzy autokorelacji wejść służy sprawdzeniu, czy ciągi wejściowe spełniają wymagania odpowiedniego pobudzenia. Przede wszystkim poszczególne sygnały wejściowe powinny być ciągami nieustannie pobudzającymi stopnia $dA + dF_k + dB_k$. Jednak w przypadku obiektu *MISO* ciągi wejściowe powinny być także wzajemnie nieskorelowane, gdyż w przeciwnym przypadku prawidłowa identyfikacja poszczególnych torów sterowania byłaby niemożliwa. Stąd bardzo ważne jest badanie stopnia uwarunkowania macierzy korelacji własnych i wzajemnych dla wszystkich wejść, a jeżeli okaże

się, że macierz ta jest słabo uwarunkowana, to badanie odpowiednich macierzy dla poszczególnych par sygnałów wejściowych w celu określenia, która para stanowi przyczynę słabego uwarunkowania, por. dodatek B.

W systemie *MULTI-EDIP* testy pomocnicze wykonywane są automatycznie, w przypadku gdy wykryto słabe uwarunkowania równań estymatora. Dzięki takiemu rozwiązaniu użytkownik nawet bez dużego doświadczenia może łatwo odkryć prawdopodobne przyczyny tego stanu rzeczy.

Oprócz wymienionych wyżej testów uwarunkowania odpowiednich macierzy, w poszukiwaniu przyczyn słabego uwarunkowania macierzy danych pomiarowych mogą pomóc testy *skracania zer i biegunów* oraz *korelacji wzajemnych wejść z błędami predykcji*. Ten pierwszy umożliwia wykrycie nadparametryzacji w modelu, natomiast ten drugi służy wykryciu obecności sprzężenia zwrotnego między wyjściem a wejściem obiektu w trakcie trwania doświadczenia identyfikacyjnego, por. [101].

Rozdział 6

Implementacja systemu komputerowego wspomagania identyfikacji

W rozdziałach 3 ÷ 5 przedstawiono dyskusję zadań wykonywanych w ramach procedury identyfikacji, ze szczególnym zwróceniem uwagi na rozgraniczenie, co może być wykonywane przez komputer, a za co odpowiedzialność musi wziąć człowiek. W komputerowo wspomaganej identyfikacji współpracują bowiem w sposób *interaktywny* [17]:

- *osoba prowadząca identyfikację*, dostarczająca *założeń* i *hipotez* na temat źródła danych (**R**, **S**), które mogą pochodzić z dwóch źródeł: dostępnej *informacji a priori* oraz nabytego już *doświadczenia* w tym zakresie,
- *komputer*, który wykonuje ściśle zdefiniowane zadania przetwarzając dane surowe i wyniki pośrednie.

Zadanie osoby prowadzącej identyfikację polega głównie na:

- podawaniu sekwencji poleceń wykonywanych przez komputer w postaci wywoływania odpowiednich *narzędzi do identyfikacji*;
- wprowadzaniu informacji *a priori* wyrażonej jako *struktura* modelu oraz jako *wartości parametrów* potrzebnych do rozwiązania zadania.

Z pierwszym rodzajem działań związane jest założenie o *domniemanej akceptacji* wstępnych ograniczeń, które są narzucone przez każde wybrane narzędzie, a co za tym idzie, uzależnienie poprawności wyników od tych ograniczeń. Komputer w zasadzie nie jest w stanie określić pewnych warunków wstępnych (założeń) i dlatego może przedstawić błędne wyniki bez ostrzeżenia, por. 5.2.2. W szczególności odnosi się to do założeń związanych z warunkami, w jakich dane doświadczalne zostały uzyskane. Wartości parametrów i struktura modelu są na ogół łatwiejsze do sprawdzenia na drodze testowania i dlatego ten rodzaj informacji *a priori* stanowi mniejsze zagrożenie jako potencjalne źródło błędów.

Z wyróżnionych w 5.2 zadań w procedurze identyfikacji planowanie eksperymentu **R** nie stanowi zadania czysto matematycznego i dlatego musi być wykonane, przynajmniej częściowo, przez człowieka. System wspomagania identyfikacji może tu jednak dostarczać pewnej *pomocy*, wyświetlając np. odpowiednie informacje o warunkach prawidłowego

wykonania doświadczenia lub dostarczając narzędzi do jego wykonania, por. rozdział 3. Podobnie struktury modelu $\mathcal{M}$ nie da się, przynajmniej obecnie, wywnioskować na podstawie samych danych i stąd komputer nie jest w stanie wykonać tego zadania samodzielnie. Co najwyżej może tu przedstawić założenia istotne przy wyborze formy $\mathcal{F}$ potrzebnej do określenia struktury $\mathcal{M}$, które osoba prowadząca identyfikację powinna *zatwierdzić*. W ten sposób to ona przejmuje za nie *odpowiedzialność*, jako że tylko ona dysponuje taką informacją, której nie ma komputer. W komputerze może być natomiast zaprogramowana pętla wewnętrzna w procedurze identyfikacji, umożliwiająca określenie zbioru wskaźników struktury μm dla założonego ciągu struktur hierarchicznych $\{\mathcal{M}(m)\}$, por. 5.2.

Z kolei zadania wyznaczenia modelu $\hat{\mathcal{M}}$ na drodze przetwarzania zbioru danych $(\mathbf{U}_N, \mathbf{Y}_N)$ dla założonej struktury $\mathcal{M}$ i kryterium aproksymacji $\mathcal{A}$ oraz jego testowanie są już zadaniami czysto matematycznymi, które mogą być wykonane przez komputer, przy założeniu że znane są odpowiednie algorytmy ich realizacji i są one numerycznie rozwiązywalne. W przeciwnym przypadku osoba prowadząca identyfikację może być zmuszona do współudziału także w tej fazie identyfikacji.

Tak więc z wymienionych w 2.1.3 podstawowych umiejętności, które potrzebne są w trakcie realizacji procedury identyfikacji, niektóre z nich mogą być przedstawione w sposób bardziej sformalizowany, szczególnie te, które związane są z zadaniami czysto matematycznymi. Umożliwia to zaprogramowanie pewnych procedur, które inteligentnie wspomagałyby osobę prowadzącą identyfikację. Typowym przykładem jest tu zaproponowany w 5.3.2 algorytm poszukiwania struktury modelu liniowego. Tam jednak, gdzie jest to trudne czy wręcz niemożliwe, wspomaganie takie może polegać po prostu na przedstawieniu *podpowiedzi* sugerujących pewne rozwiązania. Podpowiedzi takie powinny bazować *na wiedzy* i *doświadczeniu projektanta* systemu komputerowego wspomagania identyfikacji, pozostawiając *końcową decyzję użytkownikowi*, który kierując się konkretnym zastosowaniem systemu przejmuje za nią odpowiedzialność.

Złożoność programu wspomagania identyfikacji zależeć będzie więc od tego, jak ogólne struktury $\mathcal{M}$ program ten będzie obsługiwał oraz jakie metody *aproksymacji* i *testowania* zostaną w nim zaimplementowane, a także w jakim stopniu będzie on wypełniał zadanie prowadzenia użytkownika przez meandry identyfikacji. Nie ulega jednak wątpliwości, że użyteczność takiego programu będzie zależała także w istotny sposób od jakości *interfejsu użytkownika*. Dlatego w punkcie 6.1 przedstawiono ogólne wymagania użytkowe i projektowe wobec oprogramowania dla komputerowego systemu wspomagania identyfikacji. Natomiast w 6.2 omówiono pokrótce najważniejsze cechy systemu *MULTI-EDIP*, który został zaprojektowany zgodnie z przedstawionymi wymaganiami.

6.1. Oprogramowanie systemu

Można zaryzykować twierdzenie, że implementacja systemu komputerowego wspomagania identyfikacji zakończy się sukcesem nie tyle wtedy, gdy powstanie narzędzie, w którym zaprogramowano odpowiednio wyrafinowane procedury numeryczne, ale raczej wtedy, gdy użytkownik będzie w stanie korzystać z niego bez trudności, a przede wszystkim w ogóle będzie *chciał* je stosować. Jednym z podstawowych warunków sukcesu jest w tym względzie korzystanie z bogatych doświadczeń projektantów innych aplikacji. Trudno jest tu oczywiście przedstawić wyczerpujący zestaw wymagań czy zaleceń dla projektanta takiego systemu, zwłaszcza że trzeba by sięgnąć do psychologii czy ergonomii, co oczywiście wykracza poza ramy tej pracy.

Obecnie zdecydowana większość aplikacji tworzona jest dla powszechnie stosowanego systemu Windows. Jedną z zasadniczych zalet korzystania ze środowiska Windows jest standardowy, jednolity wygląd i sposób korzystania z aplikacji. Jeżeli użytkownik ma już doświadczenie w tym względzie, to może łatwo poradzić sobie z nową aplikacją. Pod warunkiem oczywiście, że będzie ona spełniała pewne minimum wymagań postulowanych przez firmę Microsoft. Przede wszystkim dotyczy to zasad tworzenia graficznego interfejsu użytkownika *GUI* (ang. *Graphical User Interface*), który został zaprojektowany właśnie w tym celu, aby użytkownik nie musiał za każdym razem rozpoczynać poznawania zasad korzystania z aplikacji od początku, nie musiał pamiętać nazw poleceń i ich składni, a korzystanie z aplikacji sprowadzało się głównie do wskazywania myszką odpowiednich ikon lub pozycji w rozwijalnych menu.

Tak więc projektując komputerowy system wspomagania identyfikacji trzeba oczywiście uwzględnić ogólne zasady tworzenia aplikacji dla środowiska Windows, takie jak np. to, że użytkownik powinien świadomie inicjować operacje, a nie wyłącznie reagować na zdarzenia w komputerze, czy to, że powinien być informowany o bieżącym stanie czy trybie pracy, a nie musiał sam domyślać się takich rzeczy, por. np. [109]. Do najważniejszych aspektów programowania aplikacji dla środowiska Windows należy także spójność, zarówno wewnętrzna, jak i z innymi aplikacjami. Jeżeli aplikacja będzie współpracować z użytkownikiem zgodnie z ogólnymi zasadami przyjętymi w Windows, to użytkownik będzie mógł poświęcić cały swój czas na rozwiązywanie interesujących go problemów, a nie na zgłębianie tajników użytkowania programu. Jedną z podstaw sukcesu w użytkowaniu programu jest korzystanie z techniki skojarzeń i porównań, a więc aplikacja powinna być tak zaprojektowana, aby korzystanie z niej było intuicyjnie oczywiste. Jednocześnie powinna być ona łatwa do przyswojenia i użytkowania, stąd potrzebny jest pewien kompromis pomiędzy elastycznością a prostotą – im

więcej opcji, tym większa elastyczność, ale większa możliwość pogubienia się użytkownika. Podobnie potrzebny jest kompromis między zasadą udostępniania użytkownikowi jak największej ilości informacji a prostotą i przejrzystością. A ponieważ dzisiaj często użytkownicy pomijają dokładne zapoznanie się z instrukcją obsługi aplikacji i od razu usiłują z niej korzystać robiąc to metodą prób i błędów, więc aby móc poznawać aplikację na drodze eksperymentowania z nią, użytkownik powinien mieć możliwość nie tylko wypróbowania każdej opcji, ale także wycofania się z niej. No i oczywiście estetyka - aplikacja powinna być tak zaprojektowana, aby była przyjemna dla oka. Dotyczy to zarówno kolorów, jak i rozmieszczenia elementów sterujących – od ich ułożenia zależy łatwość posługiwania się aplikacją.

6.1.1. Interfejs użytkownika

Ponieważ użytkownik komunikuje się z aplikacją poprzez tzw. interfejs użytkownika, stąd nie do przecenienia dla wartości danej aplikacji jest odpowiednie zaprojektowanie tego interfejsu. Decyduje on często o pierwszych wrażeniach i odczuciach osoby korzystającej z programu. Jednocześnie trzeba się liczyć z tym, że wielu użytkowników zapoznaje się z technikami identyfikacji za pomocą dostępnego oprogramowania, bez uprzedniego zaznajomienia się z odpowiednimi podręcznikami [80]. Interfejs użytkownika będzie pełnił więc dla wielu z nich istotną rolę w *poznawaniu teorii i metodyki* identyfikacji. Kierując się więc ogólnymi zasadami tworzenia aplikacji trzeba jednocześnie wziąć pod uwagę specyfikę problemu identyfikacji.

Ponieważ identyfikacja jest zadaniem skomplikowanym, wieloaspektowym, służącym wielu różnym celom, więc lista różnych możliwości, struktur modeli czy zmiennych projektowych jest bardzo długa, a dla nowicjusza może być przytłaczająca. Tak więc istota problemu projektowania oprogramowania do identyfikacji polega na tym, aby użytkownik był świadomy, jakie opcje czy wybory ma do dyspozycji, a jednocześnie nie został nimi przytłoczony. Chodzi o to, aby użytkownik niedoświadczony na początku nie czuł się całkowicie zagubiony, a jednocześnie mógł po pewnym czasie zorientować się w różnych możliwościach systemu. Najprostsze rozwiązanie polega na tym, by pokazywać pewne istotne opcje, a ukrywać inne, których użytkownik nie potrzebuje lub o których nie chce wiedzieć. W szczególności wartości wszystkich takich parametrów, które są niezależne od danych, a które dobierane są w sposób heurystyczny, powinny być wprowadzone przez projektanta.

Najłatwiej można to zrobić wykorzystując własności środowiska *GUI*, w szczególności tworząc hierarchiczny system *zagnieżdżonych* menu oraz system *kontekstowych menu* (ang. *pop-up menu*), które pokazują tylko opcje dostępne na danym etapie. Bardziej zaawansowane zmienne projektowe mogą być ukryte głębiej, w zagnieżdżonych dialogach. Tu

oczywiście pojawia się problem kompromisu pomiędzy odwoływaniem się do potrzeb użytkownika początkującego, który potrzebuje zrozumienia problemu, a łatwością i wygodą użytkowania dla osoby już zaawansowanej w identyfikacji modeli. Obie własności są potrzebne, ale stoją w pewnej sprzeczności, stąd konieczne jest pewne doświadczenie w celu osiągnięcia jakiegoś „złotego środka".

Ponieważ nie można wszystkich opcji pokazywać użytkownikowi, więc standardowa praktyka polega na ukrywaniu typowych wartości domyślnych, które są niezależne od danych, jak np. maksymalna liczba iteracji lub warunek zakończenia procedury iteracyjnej, niektóre wartości startowe dla procedur poszukiwania, wartości graniczne dla określenia, czy zadanie jest dobrze czy źle uwarunkowane, sposoby obliczania współczynników kontrakcji czy zapominania w algorytmie *RPE*, por. A.2, itp. Podstawowa trudność przy ich wyborze polega na tym, że często nie bardzo wiadomo, jaka wartość byłaby optymalna i dobiera się je na zasadzie heurystycznej. Stąd pytanie przeciętnego użytkownika o tego typu parametry byłoby dla niego niepotrzebnie kłopotliwe, nie mówiąc już o niebezpieczeństwie zaprowadzenia go na manowce. Inna sytuacja występuje wtedy, gdy dany parametr zależy od własności danych albo celu modelowania, jak np. współczynnik zapominania dla aproksymacji modeli niestacjonarnych, wymagana rozdzielczość dla modeli częstotliwościowych, kryterium wyboru struktury modelu itp. Deklaracja tego typu parametrów powinna być zostawiona użytkownikowi, co najwyżej można sugerować mu pewne podpowiedzi.

Drugie ważne pytanie to czy interfejs powinien być *zorientowany na procedurę* czy *zorientowany na wynik*, por. [80]. W pierwszym przypadku interfejs jest zorientowany na pokazanie *kolejności* działań w ramach procedury identyfikacji, a więc: generowanie danych, przygotowanie danych, wybór struktury modelu itd. Takie podejście, reprezentowane przez system *EDIP* [98], ma zdecydowaną przewagę dla użytkownika początkującego. Druga możliwość to skupienie się na jawnym przedstawianiu zmodyfikowanych zbiorów danych i różnych modeli, które otrzymano w trakcie sesji. Jest to ważne dla użytkownika doświadczonego, dla którego najważniejszym problemem jest dokumentowanie swoich działań i wyników. Dla użytkownika mniej doświadczonego takie rozwiązanie może się również okazać przydatne, gdyż pokazuje, czy zadanie identyfikacji można uznać za zakończone. W przedstawionym w punkcie 6.2 systemie *MULTI-EDIP* zastosowano to drugie podejście.

Użytkownik powinien posiadać także pewną umiejętność łączenia szeregu różnych procedur numerycznych i graficznych w jedno *użyteczne i efektywne narzędzie* do identyfikacji. Dotyczy to szczególnie takich pakietów programowych, z których korzystanie polega na *pisaniu instrukcji* wywołujących odpowiednie funkcje numeryczne lub graficzne. W tym celu konieczna jest zarówno znajomość składni poleceń dla użytkowanego systemu, jak i duża

wiedza na temat identyfikacji. Inną możliwość stanowi tu zaprojektowanie takiego systemu, który będzie użytkownika prowadził krok po kroku przez procedurę identyfikacji. Zagadnienie to jest tematem następnego podpunktu.

6.1.2. Elastyczność systemu a prostota użytkowania

Podstawowy problem, jaki staje przed projektantem systemu komputerowego wspomagania identyfikacji, to jak dostosować system do umiejętności i wiedzy użytkownika. Można rozpatrzyć tu dwa skrajne przypadki. Użytkownik o małym doświadczeniu będzie oczekiwał, że system poprowadzi go niejako „za rękę" przez kolejne etapy procedury identyfikacji, podczas gdy użytkownik o dużym doświadczeniu będzie wolał raczej samemu eksperymentować i poszukiwać własnych rozwiązań.

W pierwszym przypadku struktura systemu powinna być raczej sztywna, a wszystkie procedury numeryczne czy graficzne zaimplementowane w systemie. Zakłada się tu, że wiedza dziedzinowa użytkownika na temat identyfikacji jest niewielka, stąd ilość dostarczanej mu informacji powinna być tylko taka, jaka jest *niezbędna* do zrozumienia tego, co jest wykonywane, lub do podejmowania indywidualnych decyzji. Na każdym etapie identyfikacji użytkownikowi przedstawiane są tylko te opcje, które są istotne na tym etapie, a użytkownik co najwyżej dokonuje wyboru spośród nich. Liczba parametrów, które może wprowadzać, jest ograniczona do niezbędnego minimum, a wszystkie działania, które daje się zautomatyzować, powinny być zaprogramowane w systemie. Aby jednak zapewnić minimum elastyczności, użytkownik musi mieć możliwość dokonywania pewnych wyborów, a w szczególności *prawa veta* lub *prawa zmiany* w odniesieniu do wszystkich porad ekspertowych udzielanych przez system, jeśli uzna to za stosowne. Użytkowanie systemu przypomina tu więc korzystanie z *inteligentnego* przyrządu mikroprocesorowego, tzn. polega głównie na wyborze licznych opcji w ciągach *menu* przedstawianych na ekranie monitora. Jest to przypadek interfejsu użytkownika ukierunkowanego na procedurę identyfikacji. Takie podejście do zadania wspomagania identyfikacji reprezentowane było w wersji systemu *EDIP* dla środowiska *DOS*, por. [98].

Dla użytkowników nie będących ekspertami w dziedzinie identyfikacji zaproponowane rozwiązanie jest na pewno bardzo bezpieczne i wygodne, gdyż umożliwia im korzystanie z osiągnięć teorii identyfikacji bez konieczności dokładnego wnikania w tajniki tej dziedziny. Jednak dla kogoś, kto ma już pewne doświadczenie w tym zakresie, korzystanie z takiego systemu może być nużące. Wymaga bowiem ciągłego przechodzenia przez te same ścieżki i odpowiadania na te same pytania. Ponadto system taki jest mało elastyczny, gdyż nie umożliwia testowania własnych procedur identyfikacji. Tak więc mimo że dla badaczy

zaawansowanych w technikach identyfikacji system taki może być również bardzo użyteczny, to jednak fakt, że jest to system zamknięty, może jednocześnie stanowić istotne ograniczenie.

Na drugim biegunie mamy sytuację, gdy użytkownik chce samodzielnie testować *własne procedury identyfikacji*, tworzyć własne algorytmy lub modyfikować istniejące. Szuka on wtedy systemu, w którym będzie miał do dyspozycji odpowiednio bogaty zestaw funkcji, za pomocą których będzie mógł w *dowolny sposób* przetwarzać dane i prezentować wyniki. Klasycznym przykładem takiego rozwiązania są np. systemy *MATLAB* czy *Octave*. Korzystanie z tych systemów polega na *programowaniu* obliczeń matematycznych, które stanowią algorytm przetwarzania danych odpowiadający wybranej metodzie identyfikacji. Programowanie to odbywa się przy użyciu instrukcji oraz bibliotek funkcji dostępnych w danym środowisku lub w dołączonych przybornikach (ang. *toolbox*). W przypadku środowiska *MATLAB* są to np. *System Identification Toolbox* [81], *CONTSID Toolbox* [36], *Frequency Domain System Identification Toolbox* [68], *Adaplab-m* [3] i inne. Zapewnia to użytkownikowi olbrzymie możliwości wykorzystania gotowych funkcji, jak i tworzenia własnych. Wymaga jednak *znajomości składni* języka, a także wiedzy, jakie w ogóle instrukcje ma do dyspozycji. Dla użytkownika nie związanego ze środowiskiem akademickim, albo pracującego w takim środowisku, ale w kierunkach dalekich od automatyki lub nauk technicznych, może stanowić to istotną trudność. Przede wszystkim wymaga jednak *dobrego poznania* zagadnień związanych z dziedziną, w której pakiet jest używany – w tym przypadku z identyfikacją procesów. Użytkownik musi po prostu wiedzieć, jak należy postępować, jaka powinna być kolejność działań, co na danym etapie można zrobić, a co jest niedopuszczalne. Stąd użytkownik takiego systemu *musi być* poniekąd także *ekspertem* w zakresie identyfikacji. Wprawdzie wymienione wyżej systemy są bardzo elastyczne i pozwalają zaoszczędzić wiele pracy – nie trzeba samodzielnie pisać wszystkich procedur numerycznych czy graficznych – ale oferowany przez nie zakres inteligentnego wspierania użytkownika w identyfikacji procesów jest na ogół bardzo ubogi, chociaż i tu widać zmiany. Na przykład, w nowej wersji przybornika *System Identification Toolbox* wprowadzona została funkcja *advice*, której celem jest dostarczenie użytkownikowi pewnych porad czy komentarzy istotnych na danym etapie identyfikacji, por. [81].

W nowej wersji systemu *MULTI-EDIP*, por. punkt 6.2, zaproponowano rozwiązanie stanowiące pewien kompromis pomiędzy prostotą użytkowania a elastycznością systemu. W systemie tym także zaimplementowane są wszystkie procedury numeryczne oraz graficzne, przez co użytkownik nie ma potrzeby (i możliwości) samodzielnego programowania. Przyjęto natomiast, że użytkownik może w dowolny sposób organizować przebieg procedury identyfikacji. Oznacza to, że to on określa kolejność działań, dzięki czemu może lepiej

dopasować tę procedurę do własnych potrzeb i przyspieszyć uzyskanie modelu. W skrajnym przypadku od odczytania danych z pliku do przedstawienia modelu wystarczy kilka ruchów myszką. Tu interfejs użytkownika zorientowany jest na prezentację wyników, por. 6.1.1.

6.1.3. Dwa paradygmaty programowania

Przedstawiona w [100] definicja systemu inteligentnego wspomagania odwołuje się do *usług* świadczonych przez oprogramowanie, a nie do sposobu *technicznej realizacji* tych usług. W zasadzie można wyróżnić dwa różne podejścia do zadania projektowania systemu inteligentnego wspomagania identyfikacji: podejście konwencjonalne, oparte na opisie całości zagadnienia za pomocą *drzewa decyzyjnego* i podejście bazujące na *ortodoksyjnym systemie eksperckim*, w którym wiedza dziedzinowa zapisana w postaci reguł produkcyjnych stanowiących tzw. *bazę wiedzy* jest przedmiotem wnioskowania logicznego, realizowanego przez *system wnioskujący*, stanowiący oddzielną część programu.

W publikacjach dotyczących tej tematyki, szczególnie w końcu lat osiemdziesiątych, np. [37, 40, 45, 75, 88, 92], autorzy najczęściej milcząco sugerowali, że ścisłe podejście eksperckie jest *jedynym* właściwym podejściem do problemu inteligentnego wspomagania identyfikacji. Tymczasem, jak pokazano w [100], gdzie dokonano porównania obu rozwiązań dla identyfikacji modelu typu *ARX*, każde z nich może prowadzić do identycznie działającego oprogramowania. Ta sama *wiedza* może bowiem istnieć nie tylko w *bazie wiedzy*, ale również w *gałęziach drzewa decyzyjnego*. Żaden z wymienionych paradygmatów programowania nie ma decydującej przewagi nad drugim w sensie kosztów i nakładów pracy.

Ważnym elementem warunkującym wybór są *preferencje programisty*. Drzewo decyzyjne jest z pewnością najlepszym punktem wyjścia, niezależnie od tego, czy przewiduje się rozwiązanie konwencjonalne czy eksperckie. Przede wszystkim zauważenie potrzeby wprowadzenia dodatkowego bloku decyzyjnego w schemacie blokowym jest łatwiejsze niż zauważenie potrzeby nowej reguły w liście wielu trudno czytelnych reguł. Dzięki temu rozwiązanie konwencjonalne może stanowić podejście bardziej racjonalne (w sensie nakładów i efektów) do problemu, w szczególności gdy stosowana wiedza jest raczej ustalona i jej zaprogramowanie nie będzie wymagać zmian. Stosowanie paradygmatu właściwego dla systemów eksperckich ma niewątpliwie zalety w sytuacji *prototypowej*, gdy nie można być całkowicie pewnym kompletności i spójności wiedzy i przewiduje się jej częste zmiany. W takiej sytuacji bowiem każdorazowa zmiana wiedzy zawartej w programie konwencjonalnym jest bardziej skomplikowana pod względem programowym. Jednak na etapie opracowywania wersji finalnej systemu inteligentnego wspomagania wymogi efektywności numerycznej zdecydowanie preferują rozwiązanie konwencjonalne.

Z prac nad tworzeniem inteligentnego systemu wspomagania identyfikacji wynika także, że niekompletność czy niespójność istniejącej wiedzy dziedzinowej i konieczność jej uzupełniania o działania, które jest trudno sformalizować (np. wizualna ocena przebiegów czasowych), jest znacznie większą przeszkodą w tworzeniu *inteligentnych programów*, aniżeli ograniczenia takiego czy innego paradygmatu programowania. Ponadto równoczesne stosowanie obu podejść, tzn. równoległe tworzenie drzewa decyzyjnego i bazy wiedzy, jest nieocenione dla wykrywania luk i niespójności w posiadanej wiedzy, szczególnie na wczesnym etapie projektowania, bez względu na to, w jakim paradygmacie ostatecznie system zostanie zrealizowany. A poza tym, program wspierający nie jest *inteligentny* dlatego, że został skonstruowany na podstawie paradygmatu ortodoksyjnego systemu eksperckiego, ale dlatego, że *oferuje usługi*, które odpowiadają usługom oferowanym przez dobrego *eksperta*.

Uwzględniając powyższe uwagi, przy projektowaniu komputerowego systemu wspomagania identyfikacji *MULTI-EDIP* wybrano podejście konwencjonalne. Nie bez znaczenia był tu fakt, że oprogramowanie do identyfikacji musi zawierać mnóstwo różnego rodzaju operacji numerycznych, które w przypadku systemów reprezentujących podejście eksperckie są znacznie trudniejsze do zaprogramowania, a ich wykonywanie jest bardzo nieefektywne. Ponadto tworzenie obecnie aplikacji dla środowiska *Windows* jest ułatwione dzięki wprowadzeniu technologii *RAD* (*Rapid Application Development*) i dostępności do systemów programowania stanowiących zintegrowane środowisko programisty (*Integrated Development Environment*, *IDE*), takich jak np. *Delphi*, *Visual Basic* i inne.

6.2. System MULTI-EDIP

System *MULTI-EDIP* jest zintegrowanym środowiskiem programowo-sprzętowym przeznaczonym do komputerowego wspomagania identyfikacji, analizy i symulacji modeli matematycznych sygnałów wektorowych lub obiektów o wielu wejściach i wielu wyjściach. Podstawowym elementem systemu jest przyjazny dla użytkownika pakiet programowy o tej samej nazwie, który prowadzi użytkownika przez wszystkie działania składające się na identyfikację procesów, poczynając od planowania eksperymentu, a kończąc na testowaniu modeli. System ten dostarcza najważniejszych narzędzi w postaci zaprogramowanych procedur numerycznych i procedur graficznych, tworzących całość prostą w użytkowaniu.

Wspomaganie użytkownika w systemie obejmuje następujące ogólne i szczegółowe problemy identyfikacji:

- przedstawianie właściwej kolejności działań;

- przedstawianie właściwych zestawów opcji we wszystkich sytuacjach, które tego wymagają;
- testowanie stacjonarności procesów oraz poprawności danych;
- wybór metody identyfikacji;
- wybór struktury modelu;
- testowanie stopnia uwarunkowania w zadaniu estymacji parametrów modelu;
- testowanie struktury i parametrów modelu.

Bardzo ważną własnością systemu *MULTI-EDIP* jest to, że użytkownik nie musi posiadać żadnej wiedzy i żadnych umiejętności z dziedziny programowania komputerów, gdyż użytkuje go tak jak *przyrząd mikrokomputerowy*, tzn. wybierając rozliczne opcje w ciągach *menu* przedstawianych na ekranie monitora.

6.2.1. Historia systemu

System do komputerowego wspomagania identyfikacji procesów rozwijany był w Instytucie Automatyki Politechniki Śląskiej w latach 1985-93 pod nazwą *EDIP*, por. [98], oraz w wersji angielskojęzycznej jako *EFPI*, por. [97]. Przeznaczony był on do identyfikacji skalarnych ciągów czasowych oraz obiektów o jednym wejściu sterującym i jednym wyjściu dostępnym pomiarowo.

Doświadczenia zebrane w trakcie wdrażania systemu *EDIP* u licznych użytkowników, zarówno w uczelniach technicznych, jak i w przemyśle oraz wojsku, wskazywały na potrzebę jego rozszerzenia na wektorowe ciągi czasowe i obiekty o wielu wejściach i wielu wyjściach, co zaowocowało stworzeniem w latach 1993-97 systemu pod nazwą *MULTI-EDIP*, por. [101].

Obie wymienione wersje pracowały w środowisku systemu operacyjnego *DOS*. W związku z coraz powszechniejszym stosowaniem systemu *Windows*, oferującym znacznie wygodniejsze środowisko graficzne niż *DOS*, opracowana została nowa wersja pakietu *MULTI-EDIP*. Składają się na nią trzy aplikacje: *MULTI-EDIP*, *SYMULATOR* i *PAIO*, które mogą pracować pod kontrolą systemów *Windows 95* i następnych. Przy tworzeniu pakietu wykorzystano zintegrowane środowisko programisty *Delphi* [135].

Wieloletnie doświadczenia w stosowaniu systemów *EDIP* oraz *MULTI-EDIP* w dydaktyce oraz w wielu pracach badawczych umożliwiły określenie tych wszystkich założeń konstrukcyjnych i stosowanych mechanizmów, które pomyślnie oparły się zębowi czasu i w pełni zasługiwały na implementację również w wersji systemu dla środowiska *Windows*.

6.2.2. Założenia projektowe

U podstaw konstrukcyjnych systemu *EDIP* leżały następujące założenia, por. [98]:

1. Identyfikację modelu i jego analizę należy wykonać na *najwyższym poziomie abstrakcji*, tj. na poziomie *symbolicznym*.
2. Należy *uwolnić użytkownika od wszelkiego programowania* dzięki pełnemu sterowaniu wszystkimi usługami i funkcjami za pomocą systemu okien i hierarchicznych menu.
3. Należy prowadzić użytkownika poprzez meandry identyfikacji na drodze odpowiedniego *szeregowania menu*.
4. Należy użytkownikowi udzielać *porad eksperckich* we wszystkich tych problemach, w których na obecnym etapie rozwoju teorii identyfikacji jest to możliwe. Usługi te powinny umożliwić pomyślne przeprowadzenie identyfikacji użytkownikom nie będącym ekspertami w dziedzinie identyfikacji.
5. Należy pozostawić użytkownikowi *prawo veta* i *prawo zmiany* w odniesieniu do wszystkich porad eksperckich udzielanych przez system *EDIP*.

Z przedstawionych tu założeń projektowych zdecydowana większość przeszła próbę czasu i wykazała swoją zasadność. Natomiast najbardziej dyskusyjne okazało się założenie o prowadzeniu użytkownika krok po kroku przez wszystkie etapy identyfikacji, co było powodem narzekań użytkowników bardziej zaawansowanych. Problem pewnego kompromisu pomiędzy potrzebami użytkownika początkującego a zaawansowanego był poruszany przy okazji dyskusji na temat elastyczności systemu wspomagania, por. 6.1.2, oraz na temat wymagań wobec interfejsu użytkownika, por. 6.1.1.

W związku z tym w trakcie prac nad nową wersją systemu postanowiono zrezygnować z tego założenia i umożliwić użytkownikowi samodzielne określenie kolejności działań. Było to możliwe przede wszystkim dzięki stworzeniu odpowiedniego interfejsu graficznego *GUI*, por. [69]. Jednocześnie hierarchiczny system zagnieżdżonych menu oraz interakcyjny system hierarchicznej, kontekstowo-zorientowanej pomocy (ang. *help*) i podpowiedzi (ang. *hint*) nadal umożliwiają instruowanie użytkownika początkującego o tym, jak powinien postępować. Pewnym rozwiązaniem mogłoby tu być wykorzystanie koncepcji *kreatora systemu*, który uzależniałby sposób działania od parametru definiującego, czy z systemu korzysta użytkownik początkujący, który wymaga prowadzenia krok po kroku, czy użytkownik zaawansowany.

Projekt systemu *MULTI-EDIP* oparty został na koncepcji okna wielodokumentowego (*Multiple Document Interface*, *MDI*), umożliwiającego jednoczesne przedstawienie wielu wykresów zawierających różne charakterystyki modelu, a nawet wielu różnych modeli, co

znacznie ułatwia użytkownikowi analizę różnych aspektów otrzymanego modelu. Pozwala to także na porównanie własności różnych modeli uzyskanych dla tych samych danych (w tym także za pomocą różnych metod), albo porównanie modeli uzyskanych dla różnych ciągów danych pochodzących z tego samego procesu.

Jednocześnie, aby umożliwić użytkownikowi przyspieszenie pracy z systemem, wprowadzono koncepcję rozdzielenia wykonywania takich zadań, jak *generowanie* i *przygotowanie danych* czy *identyfikacja* od czynności związanych z ustalaniem *parametrów projektowych* dla tych zadań. Parametry potrzebne do działania (np. klasa modelu, jego wymiarowość, struktura, wybrane testy itd.) wprowadzane są przez użytkownika w kolejnych oknach dialogowych, stanowiących system hierarchiczny. Przed pobraniem wartości parametrów z okna sprawdzana jest ich poprawność, a w razie potrzeby wyprowadzana informacja o rodzaju popełnionego błędu lub dopuszczalnej wartości parametru. Przyjęto także zasadę blokowania możliwości ustawiania parametrów, które przy określonej konfiguracji nie miałyby wpływu na działanie systemu lub mogłyby przyjąć niedopuszczalną wartość.

Wszystkie parametry projektowe mogą zostać zapisane do tzw. *pliku konfiguracyjnego* w celu ich późniejszego wykorzystania. W ten sposób, po odczytaniu z pliku zapisanych tam wcześniej założeń i zadeklarowanych parametrów wstępnych, potrzebnych do właściwego przetwarzania danych, można od razu przystąpić do działań związanych z procedurą identyfikacji. W efekcie po odczytaniu ciągów czasowych z pliku zawierającego dane użytkownik ma możliwość uzyskania wyników w zaledwie kilku krokach, bez konieczności żmudnego przechodzenia przez kolejne menu.

Rozwiązanie takie niesie jednak ze sobą pewne niebezpieczeństwo polegające na tym, że w trakcie prowadzenia identyfikacji użytkownik nie musi w sposób jawny dokonywać pewnych wyborów czy deklarować potrzebnych wartości (może bowiem korzystać z parametrów, które zostały zapisane w pliku konfiguracyjnym), a więc może nie być w pełni świadomy podjętych wcześniej decyzji. A decyzje te oznaczają jednocześnie przyjęcie pewnych założeń, które powinny być oparte na wiedzy *a priori* o procesie i celu identyfikacji, a więc odpowiedzialność za nie spada na użytkownika systemu. Dlatego też wszystkie informacje dotyczące tych decyzji są zawsze przedstawiane w specjalnym oknie *raportu*. Okno to jest typowym oknem dokumentu tekstowego, w którym umieszczane są wszystkie informacje o podjętych działaniach i przyjętych założeniach. Ponadto użytkownik może dokonywać edycji zawartości tego okna dodając swoje komentarze.

Drugie niebezpieczeństwo wiąże się z tym, że użytkownik nie obeznany z techniką identyfikacji może spowodować całkowicie błędne działanie systemu. Jest to jednak cena, jaką trzeba było zapłacić za uelastycznienie systemu. Dlatego system *MULTI-EDIP* musiał zostać

uzupełniony o wiele dodatkowych *procedur sprawdzających*, zabezpieczających przed powstaniem przypadkowych błędów w działaniu systemu. Sprawdzanie to dotyczy zarówno poprawności zapisanych w pliku konfiguracyjnym parametrów, jak i poprawności przeprowadzanych operacji.

6.2.3. Użytkowanie systemu MULTI-EDIP

W systemie *MULTI-EDIP* przewidziano następujące tryby pracy:

- *Tryb symulacji*
  Użytkownik wykorzystując aplikację *SYMULATOR* symuluje proces (ciąg czasowy lub obiekt dynamiczny) i przechowuje wyniki symulacji w plikach systemu. Tryb symulacji znajduje zastosowanie głównie w nauczaniu i uczeniu się podstaw identyfikacji procesów oraz użytkowania systemu *MULTI-EDIP*. Może być także przydatny do analizy wyników identyfikacji procesów rzeczywistych.
- *Tryb zbierania danych*
  Wykorzystując aplikację *PAIO* oraz specjalizowany *układ wejść-wyjść analogowych* użytkownik wykonuje doświadczenie identyfikacyjne na rzeczywistym procesie i przechowuje wyniki doświadczenia w plikach systemu.
- *Tryb przygotowania danych i identyfikacji*
  Dane odczytane z plików (albo przekazane bezpośrednio z aplikacji *SYMULATOR* lub *PAIO*) są wykorzystywane w aplikacji głównej systemu *MULTI-EDIP* do przygotowania danych, identyfikacji, analizy i testowania modeli matematycznych.

Obsługa systemu odbywa się zgodnie z ogólnymi zasadami pracy w środowisku *Windows*, a więc na drodze wyboru odpowiednich opcji w systemie *zagnieżdżonych menu*, który umożliwia grupowanie poleceń w sposób *hierarchiczny*, oraz deklarowania wymaganych parametrów w systemie *hierarchicznych okienek dialogowych*.

Przed przystąpieniem do kolejnych etapów użytkownik powinien określić wszystkie *zmienne projektowe*, które będą potrzebne do pracy w trybie *przygotowania danych i identyfikacji*. Parametry te mogą być od razu w całości ustalone przez użytkownika na podstawie jego wiedzy *a priori* lub celu identyfikacji, albo wprowadzane odpowiednio przed każdą planowaną czynnością. W tym drugim przypadku mogą więc one wynikać nie tylko z wiedzy *a priori*, ale także z poprzednio wykonanych działań, np. testów sprawdzania danych. Parametry projektowe mogą również zostać odczytane z *pliku konfiguracyjnego*, jeżeli

w przeszłości przeprowadzono już sesję dla tego samego procesu i zadeklarowano wszystkie potrzebne zmienne.

Wykorzystując pozycję *Parametry* w głównym menu systemu użytkownik ma możliwość określenia:

- *wiedzy a priori* odnośnie do identyfikowanego procesu, w tym rodzaj i wymiarowość procesu (ciąg czasowy czy obiekt, co jest wejściem, a co wyjściem), stacjonarność procesu, typ i struktura modelu, zakresy dla poszukiwania struktury, a także *kryterium wyboru modelu* (*AIC*, *BIC* itd., por. 5.3.1) itp.;
- *metod identyfikacji* dla modeli nieparametrycznych (np. periodogramowa, korelacyjna) i *typów modeli* (charakterystyka amplitudowo-fazowa, odpowiedź impulsowa, funkcja autokorelacji itd.), w przypadku modeli parametrycznych metody identyfikacji są zdefiniowane poprzez wybrany typ modelu;
- *parametrów projektowych* dla wybranych metod identyfikacji w postaci liczb (np. współczynnik zapominania, liczba przesunięć dla funkcji korelacji, rozdzielczość) lub wyboru opcji (np. sposób okienkowania, dekompozycja dla pobudzeń ortogonalnych);
- *metod testowania* dla modeli parametrycznych (np. predykcja, symulacja, testy losowości itd.), czy model ma być testowany na danych sprawdzających;
- *parametrów dla przygotowania danych*, m.in. parametry filtru, sposób usuwania składowych deterministycznych itp.

Za te zmienne projektowe, które związane są z *założeniami* dotyczącymi procesu albo z *celem* identyfikacji, jak. np. wymiarowość procesu, co jest wejściem czy wyjściem itp., odpowiedzialny jest w całości użytkownik. Natomiast zmienne związane z *hipotezami*, takie jak np. struktura modelu czy stacjonarność procesu, mogą być testowane w ramach procedury identyfikacji, w związku z czym system *MULTI-EDIP* może tu wspomóc użytkownika w podjęciu decyzji co do tych zmiennych. Jako przykład może tu służyć algorytm automatycznego wyboru struktury, por. 5.3.2. Jeżeli użytkownik nie zadeklaruje jawnie zmiennych projektowych, które będą potrzebne na etapie przetwarzania, to system korzysta ze zmiennych domyślnych, ustawionych podczas poprzedniej sesji.

Kolejność podejmowanych działań w czasie pracy z systemem *MULTI-EDIP* zależy od użytkownika. Nie jest więc on prowadzony przez kolejne etapy procedury identyfikacji. Oznacza to, że może od razu przystąpić do *identyfikacji* modeli zgodnie z zadeklarowanymi zmiennymi projektowymi, albo najpierw wykonać działania związane z *przygotowaniem*

danych. Jednak może to robić także w sposób iteracyjny: najpierw w wybrany sposób przetwarzając dane, potem identyfikując model, a jeśli wynik nie będzie satysfakcjonujący, może wykonać inne przetwarzanie i ponownie próbować identyfikacji. Może też zmienić parametry projektowe dla identyfikacji, np. założyć inną strukturę modelu, metodę lub testy jakości modelu. Jeżeli w trakcie tych poczynań zostaną wykonane działania skutkujące przetworzeniem danych, jak np. wybranie podciągu danych, filtracja, usunięcie jakiejś składowej czy podobne, to system zezwala na cofnięcie ostatnio wykonanego przetwarzania i powrót do poprzednich wartości danych. Po każdej modyfikacji danych można także zapisać przetworzone dane do pliku w celu ewentualnego wykorzystania w przyszłości.

Wszelkie wyniki działań, które powinny być przedstawione w postaci *graficznej*, pojawiają się jako *okienka z wykresami* w oknie głównym aplikacji. Dotyczy to przede wszystkim przebiegów czasowych przetwarzanych danych oraz takich wyników identyfikacji, jak modele nieparametryczne czy wyniki niektórych testów. Dla każdego wykresu istnieje możliwość przeprowadzenia szeregu operacji, takich jak np. odczytywanie współrzędnych dowolnego punktu wykresu, powiększenie dowolnego fragmentu wykresu, przenoszenie wykresów pomiędzy różnymi okienkami itd., por. [61]. W przypadku identyfikacji modeli niestacjonarnych wyświetlane są wykresy *trójwymiarowe*, które obrazują *zmiany w czasie* identyfikowanych modeli. Na wykresie takim możliwe jest rysowanie i przesuwanie płaszczyzny przekroju, wyświetlanie przekroju w oknie wykresu dwuwymiarowego itp.

Natomiast wyniki identyfikacji przedstawiane w postaci *tekstowej*, takie jak otrzymane oceny parametrów modeli, wyniki niektórych testów, a także informacje o pochodzeniu danych, działaniach związanych z przygotowaniem danych czy operacjach wykonanych na plikach, są przedstawiane w specjalnym *oknie raportowym*. Zawartość tego okna może być dodatkowo edytowana w trakcie sesji, zapisana do pliku lub drukowana. W ten sposób użytkownik ma możliwość łatwego dokumentowania wykonanych przez siebie czynności.

W dodatku C pokazano przykład pracy z systemem *MULTI-EDIP* w trakcie sesji, której celem była identyfikacja modelu obiektu elektroakustycznego dla celów projektowania układu aktywnego tłumienia hałasu, a której wyniki omówiono w następnym rozdziale.

Rozdział 7

Wyniki identyfikacji obiektu elektroakustycznego

Celem tego rozdziału jest przedstawienie wyników identyfikacji dla rozpatrywanego w pracy przykładu zastosowania systemu komputerowego wspomagania do identyfikacji modeli obiektów elektroakustycznych wykorzystywanych do parametryzacji trójwymiarowego układu ATH do tworzenia lokalnych stref ciszy w pomieszczeniu opisanego w 2.3.

Zgodnie z uwagami z punktu 3.4 przeprowadzono eksperymenty identyfikacyjne przy użyciu sprzętu wykorzystanego później w czasie badań nad ATH, przy czym ciągi WSWO potrzebne do generacji sygnałów pobudzających głośniki uzyskano za pomocą systemu *MULTI-EDIP*. Następnie zebrane dane w postaci zarejestrowanych ciągów czasowych z mikrofonów wraz z ciągami pobudzającymi przekazane zostały do systemu *MULTI-EDIP*, którego inteligentne wspomaganie wykorzystano w trakcie wykonywania procedury identyfikacyjnej.

Ponieważ brak jest odpowiedniej wiedzy *a priori* do przyjęcia struktury modelu, stąd struktura ta wyznaczana była przez system wspomagania zgodnie z algorytmem przedstawionym w 5.3.2 oraz uwagami z punktu 4.3, przy czym identyfikowane były zarówno modele *MISO*, jak i *SISO* po dekompozycji wykorzystującej własności sygnałów WSWO. Otrzymane modele parametryczne porównywane były ze wzorcami uzyskanymi za pomocą estymatora *ETFE* (także z użyciem systemu *MULTI-EDIP*). Ze względu na „poszarpanie" charakterystyki amplitudowej wynikające z własności estymatora *ETFE*, por. [79], zastosowano wygładzające okno częstotliwościowe o szerokości 5 binów, por. [101]. Ponieważ omawiany system ATH zaprojektowany został do tłumienia hałasu w zakresie częstotliwości od około 40 do 140 Hz, więc przyjęto, że zaproponowane w 5.1.4 testy walidacji fazy oraz wskaźnik dokładności charakterystyki amplitudowej obliczane będą dla częstotliwości dyskretnych (wynikających z częstotliwości próbkowania f_p oraz liczby pomiarów N) w zakresie nieco szerszym, tzn. od 30 do 150 Hz.

Przedstawiono wyniki identyfikacji obiektu w sytuacji, gdy w trakcie eksperymentu do pomieszczenia nie docierał hałas, jak i w przypadku generowania hałasu o różnych własnościach.

7.1. Identyfikacja struktury

W tabeli 7.1 przedstawiono przykładowe wyniki identyfikacji struktury i osiągnięte wskaźniki jakości modelu dla jednego z obiektów elektroakustycznych. Do prezentacji wybrano wyniki identyfikacji torów wtórnych dla mikrofonu drugiego, jako że były one najtrudniejsze do interpretacji i najmniej jednoznaczne. Wykorzystano tu dwa podejścia. W pierwszym identyfikowano pełny model *MISO* o strukturze *ARX* dla dwóch kryteriów wyboru *AIC* i *BIC* (dwa pierwsze wiersze), w drugim dokonywano dekompozycji zadania identyfikacji obiektu *MISO* na niezależne zadania identyfikacji modeli SISO z wykorzystaniem specjalnej filtracji w dziedzinie częstotliwości, zgodnie z uwagami z 3.4.2 (wiersz trzeci tabeli).

Tabela 7.1

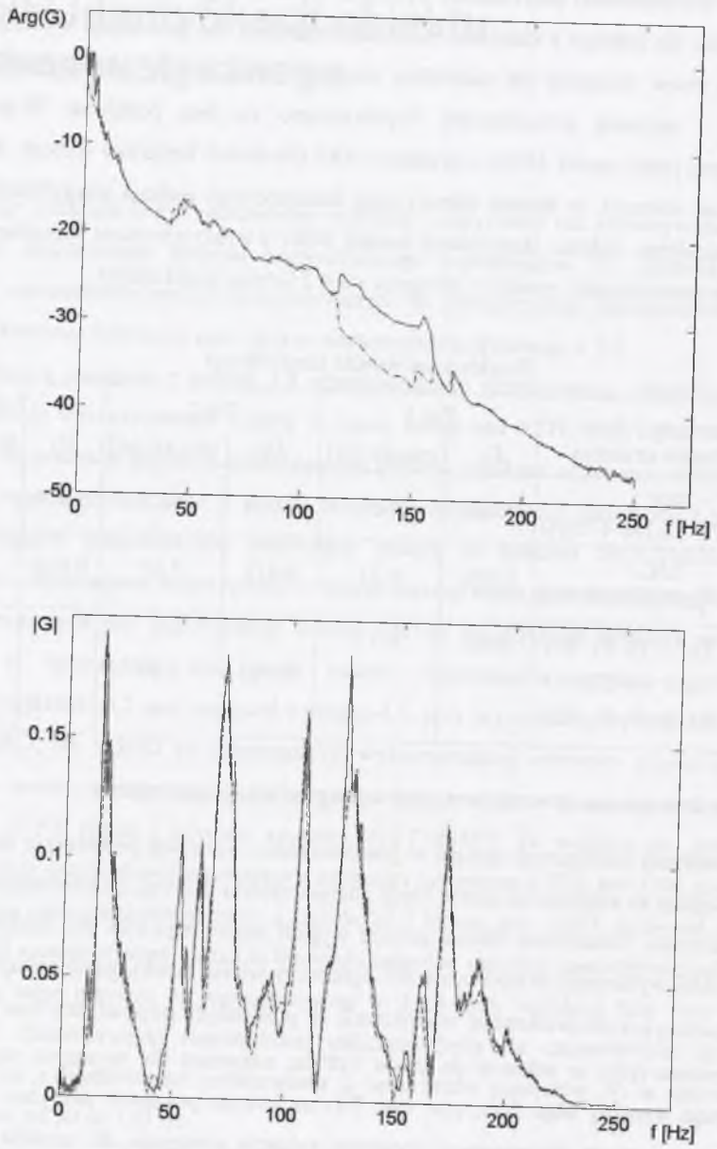
Przykładowe wyniki identyfikacji

Lp.	Kryterium wyboru / wybrana struktura	Tor 1		Tor 2		Tor 3	
		D_1	$\max\Delta\varphi_1(\omega)$	D_2	$\max\Delta\varphi_2(\omega)$	D_3	$\max\Delta\varphi_3(\omega)$
1	BIC / ([4 6 7], 51, [48 47 50])	0.026	0.25	0.014	1.25	0.019	0.74
2	AIC / ([4 6 7], 57, [55 57 50])	0.026	0.21	0.011	7.05	0.019	0.86
3	BIC / Tor 1: (4, 61, 61)	0.025	0.17	-	-	-	-
	Tor 2: (6, 53, 56)	-	-	0.010	6.90	-	-
	Tor 3: (7, 51, 50)	-	-	-	-	0.019	0.74

Na podstawie otrzymanych wyników można wyciągnąć następujące wnioski:

1. Dla testowanej konfiguracji sprzętu w pomieszczeniu występują problemy z identyfikacją toru drugiego ze względu na istotne błędy charakterystyki fazowej w interesującym zakresie częstotliwości. Stosunkowo dobrze jeszcze wypadł model tego toru dla struktury modelu *MISO-ARX* wybieranej na podstawie *BIC* (pierwszy wiersz tabeli), gdzie maksymalny błąd fazy $\max\Delta\varphi_2(\omega)$ nie przekraczał wartości $\pi/2$. W pozostałych przypadkach faza była dobrze modelowana tylko w zakresie do około 110 Hz, natomiast dla wyższych częstotliwości występuje wyraźny błąd fazy, por. rys. 7.1, na którym pokazano przykład dla modelu uzyskanego metodą dekompozycji (podobna sytuacja występuje dla modelu *MISO-ARX* wybranego dla kryterium *AIC*). Analiza charakterystyki amplitudowej pokazuje jednocześnie, że dla tej częstotliwości występuje wyraźna dolina, spowodowana prawdopodobnie wystąpieniem węzła w okolicach umieszczenia mikrofonu. Tak więc przy

takiej konfiguracji układ ATH może działać prawidłowo tylko do częstotliwości około 110 Hz.



Rys. 7.1. Porównanie charakterystyk częstotliwościowych: fazowej (po rozwinięciu) i amplitudowej modelu drugiego toru wtórnego uzyskanego metodą dekompozycji (linia przerywana) oraz metodą analizy widmowej (linia ciągła)

Fig. 7.1. Comparison of frequency responses: phase (unwrapped) and magnitude of the second secondary path model obtained by decompostion method (dashed line) and by spectral analysis (solid line)

2. W rozpatrywanym przykładzie kryterium *BIC* jest bardziej odpowiednie niż *AIC*. Modele wybrane zgodnie z *AIC* posiadają znacznie większe stopnie wielomianów, mimo że ich jakość jest w zasadzie porównywalna z modelami uzyskanymi za pomocą *BIC*, a dla toru drugiego nawet gorsza. W tym przypadku „kara" $2m$ za liczbę parametrów jest zbyt mała w porównaniu z $N \ln V(\hat{\theta})$ (por. równanie (5.11)), co w konsekwencji może prowadzić do niepotrzebnie wysokich stopni w modelu. Szczególnie dotyczy to identyfikacji modeli *SISO* z wykorzystaniem metody dekompozycji. Na przykład, identyfikacja toru G_2 przy użyciu *BIC* dostarcza wynik w postaci $\mu = (6, 53, 56)$, podczas gdy poszukiwania dla kryterium *AIC* zatrzymują się na górnym ograniczeniu równym 80. Zwiększając to ograniczenie do 120 otrzymano ponownie strukturę bliską ograniczeniu, tj. (6, 117, 116). Jednocześnie wskaźnik jakości D_2 dla struktury bogatszej nieco się pogorszył (wzrósł do wartości 0.011), a przede wszystkim pojawiły się znaczne błędy w charakterystyce fazowej dla wyższych częstotliwości. Podobnie zwiększając górne ograniczenie do 200 uzyskano strukturę (5, 191, 195), o jeszcze mniejszej wartości *AIC*, a obliczona charakterystyka fazowa była kompletnie błędna.

3. Wykorzystując możliwości, jakie daje sygnał WSWO, i dekomponując zadanie identyfikacji obiektu *MISO* na niezależne zadania identyfikacji modeli *SISO* uzyskano wyniki porównywalne z tymi, które otrzymano dla podejścia konwencjonalnego. Jednak niezaprzeczalną zaletą metody bazującej na dekompozycji jest znaczne uproszczenie zadania identyfikacji i skrócenie czasu obliczeń. W tym przypadku bowiem mamy do czynienia z trzema układami równań o wymiarach $dA_k + dB_k + 1$ każdy, zamiast jednego układu o wymiarach $dA + dB_1 + dB_2 + dB_3 + 3$, a zadanie poszukiwania struktury modelu w siedmiowymiarowej przestrzeni liczb całkowitych sprowadza się do trzech zadań poszukiwania w przestrzeniach trójwymiarowych.

Ponieważ celem identyfikacji jest znalezienie zarówno parametrów modelu, jak i jego struktury, może być interesujące prześledzenie działania algorytmu automatycznego wyznaczania struktury, który opisano w punkcie 5.3.2. Przeanalizowany zostanie przypadek identyfikacji modelu *MISO-ARX* dla kryterium wyboru *BIC* (pierwszy wiersz tabeli).

1. Oszacowanie czasu opóźnienia

Zgodnie z algorytmem w pierwszym kroku wyznacza się czasy opóźnienia w poszczególnych torach dla najniższego przyjętego stopnia modelu. W wyniku działania tego algorytmu uzyskano następujące wartości czasów opóźnienia: $\hat{d}_1 = 4$, $\hat{d}_2 = 6$ oraz $\hat{d}_3 = 7$, a więc wartości, które stanowią także wynik końcowy, co potwierdza skuteczność

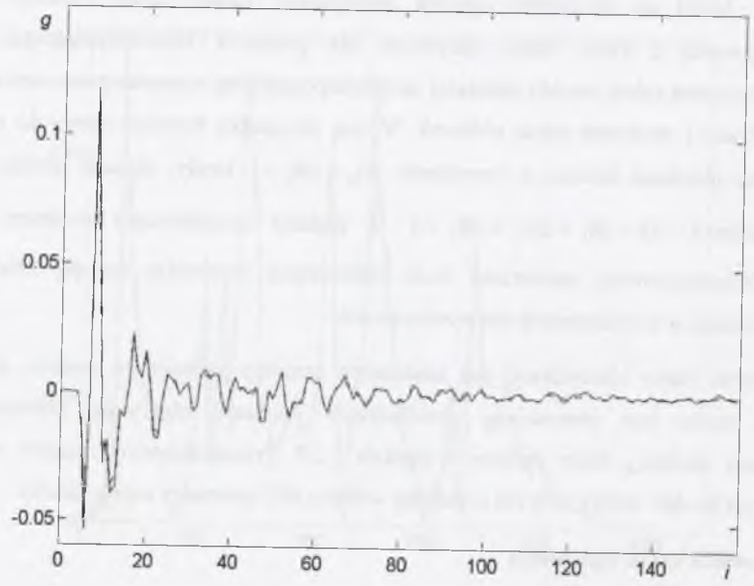
zaproponowanej metody. O tym, że są to wartości prawidłowe, świadczy porównanie przebiegu odpowiedzi impulsowej otrzymanego modelu parametrycznego z odpowiedzią impulsową uzyskaną metodą nieparametryczną dla pliku danych do walidacji. Przykład takiego porównania dla toru pierwszego dla 100 pierwszych wartości odpowiedzi pokazano na rys. 7.2.

2. Wyznaczenie punktu startowego dla poszukiwania struktury modelu MISO

Poszukiwanie struktury modeli SISO dla kolejnych torów dało następujące wyniki $(\hat{d}_1, d\hat{A}_1, d\hat{B}_1) = (5, 49, 46)$, $(\hat{d}_2, d\hat{A}_2, d\hat{B}_2) = (6, 48, 47)$ oraz $(\hat{d}_3, d\hat{A}_3, d\hat{B}_3) = (7, 49, 50)$. Dla każdego z torów procedura identyfikacji wymagała estymacji parametrów dla około 50 do 60 struktur *ARX*.

3. Poszukiwanie struktury modelu MISO

Poczynając od punktu startowego, tj. modelu o strukturze ([5 6 7], 49, [46 47 50]), po przetestowaniu 24 struktur *MISO-ARX* osiągnięto w wyniku model o strukturze ([4 6 7], 51, [48 47 50]), dla której wartość *BIC* wynosi -2.7055.



Rys. 7.2. Odpowiedź impulsowa modelu toru pierwszego uzyskanego metodą parametryczną (linia ciągła) i metodą nieparametryczną dla danych do walidacji (linia przerywana)
Fig. 7.2. Impulse response of the first secondary path model obtained by parametric method (solid line) and by non-parametric method for validation data (dashed line)

W tym miejscu warto zauważyć, że w wyniku poszukiwań nie osiągnięto minimum globalnego *BIC*. Można bowiem znaleźć inne struktury, dla których wartość *BIC* jest mniejsza, np. dla ([5 6 7], 57, [48 47 50]) *BIC* = -2.7097. Testy walidacji przeprowadzone dla tej struktury dały wyniki zbliżone do modelu otrzymanego dla kryterium *AIC* (drugi wiersz tabeli 7.1), a więc o ile modele toru pierwszego i trzeciego były porównywalne, to dla drugiego toru występował nieakceptowalnie duży błąd fazy. Tak więc w tym przypadku zakończenie procedury identyfikacji w minimum lokalnym dało paradoksalnie lepszy wynik.

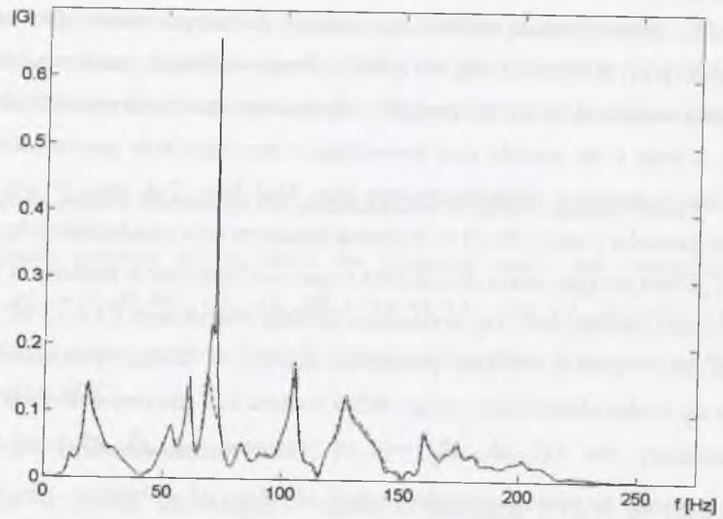
Gdyby jednak przyjąć zakres poszukiwań stopni wielomianów w modelu od 20 do 80, to wtedy procedura zakończyłaby się uzyskaniem modelu o strukturze ([5 6 7], 46, [45 31 30]) o wartości *BIC* = -2.6776. W tym przypadku niestety w trakcie poszukiwania punktu startowego na drodze identyfikacji modeli SISO wartość *BIC* dla torów drugiego i trzeciego osiąga minimum dla $(\hat{d}_2, d\hat{A}_2, d\hat{B}_2) = (6, 30, 31)$ oraz $(\hat{d}_3, d\hat{A}_3, d\hat{B}_3) = (7, 30, 30)$, co w efekcie prowadzi do minimum lokalnego dość odległego od globalnego. Przy identyfikacji tak trudnych obiektów, jak w rozpatrywanym przykładzie użytkownik musi liczyć się więc z koniecznością testowania różnych zakresów poszukiwania ze względu na występującą tu dużą zmienność wartości kryterium wyboru modelu dla kolejnych struktur hierarchicznych.

7.2. Identyfikacja w obecności zakłócenia

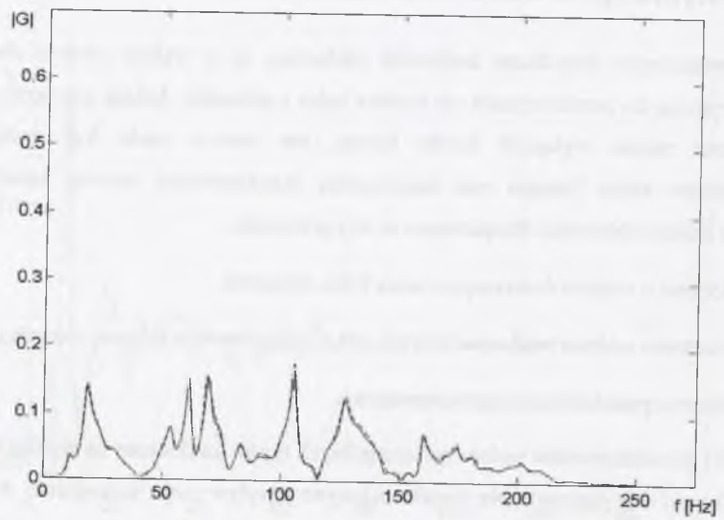
W przeprowadzonych dotychczas badaniach zakładano, że w trakcie trwania eksperymentu identyfikacyjnego do pomieszczenia nie dociera hałas z zewnątrz. Jednak założenie, że na czas eksperymentu można wyłączyć źródło hałasu, nie zawsze może być spełnione, stąd przeprowadzono także badania nad odpornością przedstawionej metody identyfikacji na dodatkowe źródło zakłócenia. Rozpatrzono tu trzy przypadki:

- zakłócenie o widmie dyskretnym (suma kilku sinusoid),
- zakłócenie o widmie wąskopasmowym, ale niestacjonarnym (sygnał pracującego silnika),
- zakłócenie pseudolosowe szerokopasmowe.

W [66] przeanalizowano wpływ poszczególnych typów zakłócenia na wyniki identyfikacji oraz możliwości ich poprawy. Na wyniki negatywny wpływ mają szczególnie dwa pierwsze typy zakłócenia, gdyż powodują istotne zafałszowanie charakterystyk częstotliwościowych modelu, por. przykład charakterystyki amplitudowej dla zakłócenia pochodzącego z pracującego silnika przedstawiony na rys. 7.3.



Rys. 7.3. Porównanie charakterystyk amplitudowych modelu uzyskanych w przypadku wystąpienia zakłócenia (linia ciągła) oraz braku zakłócenia (linia przerywana)
Fig. 7.3. Comparison of magnitude of frequency responses for models with disturbance (solid line) and without disturbance (dashed line)



Rys. 7.4. Porównanie charakterystyk amplitudowych modelu uzyskanych w przypadku wystąpienia zakłócenia i zastosowania DAF (linia ciągła) oraz braku zakłócenia (linia przerywana)
Fig. 7.4. Comparison of magnitude of frequency responses for models with disturbance and DAF (solid line) and without disturbance (dashed line)

W tym przypadku zaproponowano zastosowanie specjalnej procedury filtracji w dziedzinie częstotliwości, która umożliwiłaby usunięcie z widma sygnału wyjściowego tych częstotliwości, które odpowiadały częstotliwości zakłócenia. Metoda ta została nazwana *DAF* (ang. *Disturbance Adjusted Filtering*) i jest możliwa do zastosowania tylko w przypadku pobudzenia obiektu za pomocą sygnału wielosinusoidalnego. Przykładowy wynik zastosowania tej metody do uzyskania poprawnej charakterystyki częstotliwościowej pokazano na rys. 7.4. Przeprowadzone jednocześnie badania różnych metod identyfikacji częstotliwości, które należy odfiltrować, pokazały w tym przypadku wyższość klasycznej analizy widmowej nad metodami identyfikacji widm dyskretnych, takimi jak np. *MUSIC* czy *EV*, por. [101].

W przypadku zakłócenia pseudolosowego, dla którego moc rozkłada się równomiernie na poszczególne częstotliwości, uzyskane wyniki identyfikacji nie różniły się zasadniczo od tych, które otrzymano przy braku zakłócenia. Wyniki te można by jeszcze poprawić przez powtarzanie eksperymentu i uśrednianie, ale wiązałoby się to z koniecznością znacznego wydłużenia czasu eksperymentu.

7.3. Podsumowanie

Na podstawie przeprowadzonych badań można stwierdzić, że podejście polegające na przyjęciu kryterium aproksymacji opartego na minimalizacji błędu predykcji oraz użyciu kryterium informacyjnego jako substytutu kryterium walidacji w iteracyjnej procedurze poszukiwania struktury modelu okazało się przydatne dla postawionego celu identyfikacji. Umożliwiło bowiem uzyskanie modeli transmitancyjnych o odpowiednio dokładnych charakterystykach w dziedzinie częstotliwości. Jednocześnie kryterium *BIC*, w którym „kara" za złożoność modelu jest większa niż w *AIC*, jest bardziej przydatne w rozpatrywanym przykładzie z uwagi na stosunkowo dużą liczbę pomiarów N oraz wartość funkcji strat V.

W planowaniu eksperymentu identyfikacyjnego warto brać pod uwagę pobudzanie obiektu z wykorzystaniem sygnałów WSWO. Sygnały tego typu umożliwiają nie tylko uzyskanie lepszych wyników identyfikacji, ale także pozwalają na znaczne uproszczenie zadania identyfikacji modelu *MISO* dzięki metodzie bazującej na dekompozycji widma sygnału wyjściowego. Ponadto w przypadku istnienia zakłóceń wąskopasmowych lub o ograniczonych widmach dyskretnych umożliwiają one znaczne poprawienie uzyskanych wyników dzięki możliwości odfiltrowania z przetwarzanych sygnałów częstotliwości występujących w zakłóceniach.

Warto tu także zauważyć, że eksperymenty dodatkowe, których celem było uzyskanie charakterystyk wzorcowych dla przeprowadzenia walidacji, wykonane były jedynie w celach

badawczych, tj. sprawdzenia poprawności zaproponowanego podejścia i nie muszą być powtarzane w czasie normalnej parametryzacji układu ATH.

Jednocześnie przykład ten pokazuje dobitnie, że nie można oczekiwać od systemu komputerowego wspomagania identyfikacji pełnej automatyzacji procedury identyfikacji. Potrzebna jest, szczególnie w trudniejszych przypadkach, takich jak np. rozważany tu obiekt elektroakustyczny, wnikliwa analiza wyników. Jednak pomoc takiego systemu, jak np. *MULTI-EDIP* jest nieoceniona. Nawet w najgorszym przypadku można było uzyskać akceptowalne wyniki już po kilku, kilkunastu minutach pracy, podczas gdy uzyskanie podobnych wyników przy użyciu np. środowiska *Matlab* i *System Identification Toolbox* wymagałoby wielu godzin prac programistycznych, nie mówiąc już o tym, że wiązałoby się to także z koniecznością posiadania pewnej biegłości w ich użytkowaniu.

Korzystając z systemu *MULTI-EDIP*, mimo bardzo wysokich stopni wielomianów i dużego zakresu poszukiwań, wyniki uzyskiwano bardzo szybko dzięki odpowiedniemu zaprogramowaniu procedury najmniejszych kwadratów (por. dodatek A.1).

Przedstawiona tu metodologia postępowania w identyfikacji obiektów elektroakustycznych została wykorzystana w pracach nad tworzeniem lokalnych stref ciszy w pomieszczeniu. Badania te potwierdziły słuszność przyjętych rozwiązań. Oczywiście wyniki działania układów kompensacji, tj. rozmiary tworzonej strefy oraz stopień tłumienia hałasu, zależały nie tylko od poprawności zidentyfikowanych modeli torów wtórnego i sprzężenia zwrotnego, ale także od wielu innych czynników związanych z algorytmami sterowania i ich parametryzacją. W zależności od konfiguracji układu ATH oraz typu hałasu uzyskiwano tłumienie od kilku do kilkunastu decybeli dla zakłóceń szerokopasmowych (pseudolosowych) oraz do kilkudziesięciu dB dla zakłóceń wąskopasmowych (deterministycznych). Dokładny opis badań i uzyskanych wyników wykracza jednak poza ramy tej pracy, której celem jest dyskusja problemu komputerowego wspomagania identyfikacji. Wyniki badań nad wykorzystaniem układów ATH do tworzenia stref ciszy zostały przedstawione w pracy [90].

Rozdział 8
Wnioski końcowe

W pracy przedstawiono problem komputerowego wspomagania identyfikacji procesów. W sposób szczególny uwzględniono zadanie wspomagania w zakresie identyfikacji liniowych modeli dynamicznych typu czarna skrzynka, gdyż modele te należą do klasy najczęściej identyfikowanych i stosowanych w praktyce. Omówiono podstawowe warunki prawidłowego wnioskowania na podstawie danych zebranych z procesu oraz wymagania wobec systemu wspomagania użytkownika w skutecznym i poprawnym przeprowadzeniu procedury identyfikacji. Główny wniosek, jaki się tu nasuwa, jest taki, że automatyzacja tak skomplikowanego i wymagającego dużej wiedzy zadania jak identyfikacja procesów jest do pewnego stopnia możliwa. Jednak jest to zadanie trudne koncepcyjnie i czasochłonne, wymagające dużych umiejętności programowych, oprócz oczywiście dużej wiedzy i doświadczenia w zakresie samej identyfikacji.

Projektowanie systemu komputerowego wspomagania identyfikacji, jak każda inna działalność twórcza, stwarza nowe perspektywy, ale niesie także ze sobą możliwość powstania pewnych pułapek. Dotyczy to zarówno twórców takich systemów, jak i ich użytkowników.

Niewątpliwą korzyść, jaką odnosi *projektant* w czasie pracy nad takim systemem, jest konieczność uporządkowania własnej wiedzy, przemyślenia, co tak naprawdę jest ważne i co ma istotny wpływ na wyniki identyfikacji. Przeprowadzenie wielu badań symulacyjnych, koniecznych przy testowaniu oprogramowania, pozwala na poszerzenie wiedzy wynikającej z doświadczenia i lepsze zrozumienie problemów, jakie mogą się pojawić w trakcie identyfikacji. Nie bez znaczenia jest tu także sprzężenie zwrotne od użytkowników systemu, których uwagi czy pytania pozwalają lepiej zrozumieć ich potrzeby w zakresie identyfikacji procesów, stawiają nowe wyzwania i poszerzają horyzonty, o czym autor sam się dobitnie przekonał.

Z drugiej strony projektant staje przed pewną pokusą, aby włączyć do systemu całą swoją wiedzę, co może prowadzić do przedstawiania użytkownikowi zarówno potrzebnej, jak i zbędnej informacji. W ten sposób można wprawdzie popisywać się wiedzą, ale jednocześnie niepotrzebnie przytłoczyć użytkownika, dla którego korzystanie z takiego systemu może się stać zbyt trudne i wręcz prowadzić go na manowce. A przecież podstawowym celem projektanta systemu wspomagania identyfikacji jest pomoc użytkownikowi i satysfakcja, jaką

może on osiągnąć korzystając z systemu. Podobną pokusą może być dążenie do nasycenia tworzonej aplikacji różnymi „gadżetami", które niepotrzebnie mogą przysłonić istotę problemu.

Najważniejsze z punktu widzenia *użytkowników* systemów komputerowego wspomagania identyfikacji jest to, że osoby, które na co dzień nie zajmują się identyfikacją, mogą korzystać z doświadczeń i osiągnięć ekspertów w tej dziedzinie. Umożliwia to znaczne rozszerzenie zastosowań na wiele dziedzin nauki nie związanych bezpośrednio z automatyką, takich jak np. medycyna czy nauki biologiczne, a co za tym idzie – rozwój nauki i poszerzenie wiedzy o świecie. Istotna jest tu na pewno także oszczędność czasu i pieniędzy na badania – jest narzędzie, z którego można korzystać bez potrzeby tworzenia własnego oprogramowania. Dużą zaletą systemu może być także możliwość prowadzenia badań symulacyjnych.

W punkcie 5.2.2 przedstawiono potencjalne źródła błędów, które mogą spowodować, że identyfikacja zakończy się niepowodzeniem. Wszystkie omówione tam niebezpieczeństwa powinny być przedstawione użytkownikowi w ramach tworzonego systemu, w postaci podpowiedzi, sugestii czy ostrzeżeń. Przede wszystkim użytkownik musi zdawać sobie sprawę, że model, który otrzymał w wyniku nawet poprawnie przeprowadzonej procedury identyfikacji, nie jest tożsamy z badanym procesem, że odzwierciedla on tylko jego niektóre właściwości i pozostaje słuszny jedynie dla warunków zbliżonych do tych, w jakich było przeprowadzone doświadczenie identyfikacyjne. Musi być także świadomy, że uzyskał wynik dla konkretnej realizacji, który może silnie zależeć od losowości danych, szczególnie w przypadku słabego uwarunkowania układu równań estymatora. Istotne jest również to, by nie mylił dokładności modelu z precyzją liczb, nie mówiąc już o błędnych wnioskach, jakie mogą być efektem błędnego wykonania doświadczenia, np. wyniki mające niewiele wspólnego z rzeczywistością, gdy korelacja między sygnałami nie jest wynikiem zależności. Oczywiście, projektant systemu nie jest w stanie przewidzieć wszystkich możliwości i użytkownik powinien zdawać sobie sprawę z własnej odpowiedzialności za przyjęte *założenia* czy *niedostatecznie sprawdzone hipotezy*. To on bowiem musi sformułować problem identyfikacji, chociaż nie musi go umieć w pełni rozwiązać, ponieważ wspomaga go w tym system komputerowy.

Kolejne niebezpieczeństwo dla osoby korzystającej z systemu wspomagania identyfikacji może wynikać z zafascynowania łatwością użytkowania takiego systemu i uzyskiwania wyników, co może skutkować dążeniem do „produkowania" modeli i zastępowania w ten sposób innych działań, np. wnikliwej interpretacji wyników. W efekcie użytkownik systemu dysponuje dużą liczbą wyników, ale nie bardzo potrafi je wykorzystać i nie wie, na ile są one poprawne.

I wreszcie zagrożenie, które może czyhać szczególnie na użytkownika rozpoczynającego dopiero swoje doświadczenia z identyfikacją, to „ślepa wiara", że skoro oprogramowanie do

wspomagania identyfikacji zostało zaprojektowane przez kogoś, kto jest specjalistą w tej dziedzinie, to uzyskany za pomocą tego oprogramowania wynik musi być poprawny. Niestety, jak podkreśla wielu autorów, identyfikacja procesów jest tak samo nauką, jak i sztuką, por. np. [80]. Stąd narzędzie pozostaje narzędziem, choćby nie wiadomo jak inteligentnym, a za sposób jego wykorzystania i jakość dostarczonych danych wejściowych odpowiada użytkownik.

Zaprojektowany na podstawie przedstawionych w pracy rozważań system *MULTI-EDIP* pokazał swoją dużą przydatność w realizacji wielu projektów badawczych związanych z identyfikacją procesów. Prace nad systemem pomogły także w lepszym rozumieniu problemów identyfikacji i pozwoliły uzyskać dodatkowy wgląd w jej logikę. System ten był intensywnie testowany przez rzesze studentów, dyplomantów i doktorantów na Wydziale Automatyki Politechniki Śląskiej. Wykorzystywany był także w tak różnych ośrodkach naukowych i przemysłowych, jak np. Uniwersytet Calgary (Kanada), Wojskowy Instytut Uzbrojenia, Huta Miedzi Głogów czy wyższe uczelnie techniczne. Jednocześnie, jak pokazał omówiony w rozdziale 7 przykład wykorzystania systemu *MULTI-EDIP* do identyfikacji obiektów elektroakustycznych, nie do zastąpienia jest wnikliwa analiza wyników wynikająca ze znajomości specyfiki identyfikowanego procesu, której nie da się zaprogramować w systemie.

Literatura

1. Akaike H.: *Statistical predictor identification.* Annals of Institute of Statistical Mathematics, vol. 22, 203 – 217, 1970.
2. Akaike H.: *A new look at the statistical model identification.* IEEE Trans. on Automatic Control, Vol. 19, 716–723, 1974.
3. Alexandrov A. G., Orlov Y. F., Mikhailova L. S.: *Adaplab-m: identification and adaptation toolbox for Matlab.* Proceedings of the 13th IFAC Symposium on System Identification (SYSID 2003), 995 – 1000, Rotterdam 2003.
4. Anderson B. D. O., Moore J. B.: *Optimal Filtering.* Prentice-Hall, Englewood Cliffs, New Jersey 1979.
5. Bai E.W.: *An Optimal Two-Stage Identification Algorithm for Hammerstein-Wiener Systems.* Automatica, Vol. 34, 333 – 338, 1998.
6. Bai E.W.: *A Blind Approach to the Hammerstein - Wiener Model Identification.* Automatica, Vol. 38, 967 – 979, 2002.
7. Balasubramanian V.: *A Geometric Formulation of Occam's Razor For Inference of Parametric Distributions.* http://omega.albany.edu:8008/ignorance/balasubra.pdf, 2001.
8. Barron Y., Rissanen J. and Yu B.: *The minimum description length principle in coding and modeling.* IEEE Trans. on Information Theory, Vol. 44, 2743–2760, 1998.
9. Bartlett M. S.: *On the theoretical specifications and sampling properties of autocorrelated time series.* Journal Royal Statistical Society, Ser. B, Vol. 8, 27 - 41, 1946.
10. Beheshti S., Dahleh M. A.: *A new information theoretic approach to order estimation problem.* Proceedings of the 13th IFAC Symposium on System Identification (SYSID 2003), 791 – 796, Rotterdam 2003.
11. Bendat J. S., Piersol A. G.: *Engineering Applications of Correlation and Spectral Analysis.* Wiley, New York 1986.
12. Bilinskis I., Mikelsons A.: *Randomized Signal Processing.* Prentice Hall Int., London 1992.
13. Billah M. B., Hyndman R. J., Koehler A. B.: *Empirical Information Criteria for Time Series Forecasting Model Selection.* http://ideas.repec.org/p/msh/ebswps/2003-2.html, 2003.

14. Billings, S.A.: *Identification of nonlinear systems - a survey.* IEE Proc. Part D, Vol. 127, 272 – 285, 1980.
15. Błażej M.: *Comparison of Different Control Structures of Active Noise Control Systems.* Proceedings of the 6th Conference on Active Noise and Vibration Control Methods, paper no. 006 (CD), Cracow 2003.
16. Bohlin T.: *Identification: Practical aspects.* Systems and Control Encyclopedia (M.G. Sing, Ed.), Vol. 4, 2301 - 2307, Pergamon Press, Oxford 1987.
17. Bohlin T.: *Interactive System Identification: Prospects and Pitfalls.* Springer Verlag, Berlin 1991.
18. Bohlin T.: *Derivation of a 'designer's guide' for interactive 'grey-box' identification of nonlinear stochastic objects.* Int. J. Control, Vol. 59, No. 6, 1505 – 1524, 1994.
19. Box G. E. P., Jenkins G. M.: *Time Series Analysis forecasting and control.* Holden-Day Inc., San Francisco, CA, 1976.
20. Chen H. F., Guo L.: *Identification and Stochastic Adaptive Control.* Birkhauser 1991.
21. Claskens G., Hjort N. L.: *The Focussed Information Criterion,* Journal of the American Statistical Association, vol. 98, 900 - 916, 2003.
22. Cobb G. W.: *Introduction to Design and Analysis of Experiments.* Springer-Verlag, New York 1998.
23. De Fatta D. J., Lucas J. G., Hodgkiss W. S.: *Digital Signal Processing: A System Design Approach.* John Wiley & Sons Inc., New York 1992.
24. Diekmann K., Unbehauen H.: *Recursive identification of MIMO systems.* Preprints of 5th IFAC Symp. on Identification and Parameter Estimation, Darmstadt, 423 – 429, 1979.
25. Edwards W., Lindman H., Savage L. J.: *Bayesian statistical inference for psychological research.* Psychology Revue, Vol. 70, 193 – 242, 1963.
26. Elliot S. J.: *Signal Processing for Active Control.* Academic Press, 2001.
27. Eykhoff P.: *System Identification.* Wiley, New York 1974.
28. Figwer J.: *Identyfikacja modeli częstotliwościowych obiektów wielowejściowych.* (Praca doktorska), Instytut Automatyki, Gliwice 1992.
29. Figwer J.: *Synthesis and simulation of random processes.* Zeszyty Naukowe Pol. Śląskiej, Automatyka, z. 126, Gliwice 1999.

30. Figwer J.: *Model Identification and Update Under Operation of Active Noise Control Systems*. Proceedings of the 6th Conference on Active Noise and Vibration Control Methods, paper no. 011 (CD), Cracow 2003.

31. Figwer J., Błażej M.: *Chaos in Active Noise Control Systems*. Proceedings of 10th Int. Congress on Sound and Vibration ICSV 2003, Stockholm (CD), 2003.

32. Figwer J., Niederliński A., Kasprzyk J.: *A New Approach to the Identification of Linear Discrete-Time MISO Systems*. Archives of Control Sciences, vol. 2, no. ¾, 223 - 239, 1993.

33. Fisher R. A.: *On an absolute criterion for fitting frequency curves*. Messenger of Mathematics, vol. 41, 155 – 160, 1912.

34. Forssel U., Ljung L.: *Identification for control: Some results on optimal experiment design*. Proceedings of the 37th Conference on Decision and Control, Tampa 1998.

35. Foster L.: *Solving Rank-Deficient and Ill-posed Problems using UTV and QR Factorizations*. SIAM J. Matrix Anal. Appl., vol. 25, 582 - 600, 2004.

36. Garnier H., Mensler M.: *The CONTSID Toolbox: A Matlab Toolbox for CONtinuous-Time System IDentification*. Proceedings of the 12th IFAC Symposium on System Identification (SYSID 2000), vol. 3, 851 – 856, Santa Barbara, CA, 2000.

37. Gentil S., Barraud A., Szafnicki K.: *SEXI: An Expert Identification Package*. Automatica, vol. 26, 803 - 809, 1990.

38. Główka T.: *Identyfikacja on-line w układzie aktywnego tłumienia hałasu – nowa metoda z dziedziny widm wyższych rzędów*. Materiały XV Krajowej Konferencji Automatyki, Warszawa, s. 387 – 392, 2005.

39. Godfrey K.: *Perturbation Signals for Systems Identification*. Prentice Hall Int., Cambridge University Press, 1993.

40. Gomide F. *et al.*: *Expert System Identification: The Supervisory Approach*. Preprints of 8th IFAC Symposium On Identification and System Parameter Estimation, Bejing 1691 – 1696, 1988.

41. Goodwin G. C., Payne R. L.: *Dynamic System Identification: Experiment Design and Data Analysis*. Academic Press, New York 1977.

42. Greblicki W.: *Stochastic approximation in nonparametric identification of Hammerstein systems*. IEEE Trans. on Automatic Control, Vol. 47, 1800–1810, 2002.

43. Gustavsson I., Ljung L., Söderström T.: *Identification of Processes in Closed Loop – Identifiability and Accuracy Aspects*. Automatica, vol. 13, 59 – 75, 1977.

44. Haber R., Unbehauen H.: *Structure identification of nonlinear dynamic systems – a survey on input/uotput approaches*. Automatica, Vol. 26, 651–677, 1990.

45. Haest M., Bastin G., Gevers M., Wertz V.: *ESPION: an Expert System for System Identification*. Automatica, vol. 26, 85 - 95, 1990.

46. Hannan E. J., Deistler M.: *The Statistical Theory of Linear Systems*. Wiley, New York 1988.

47. Hansen C. H., Snyder S. D.: *Active Control of Noise and Vibration*. Cambridge University Press, 1997.

48. Isermann R.: *Practical aspects of process identification*. Automatica, Vol. 16, 575 – 587, 1980.

49. Janiszowski K.: *Identification of coefficients of a low order continuous time transfer function from discrete-time recorded measurements*. Archives of Control Sciences, Vol. 10; 31-46, 2000.

50. Janiszowski K.: *Identyfikacja modeli parametrycznych w przykładach*. Akademicka Oficyna Wydawnicza EXIT. Warszawa 2002.

51. Jaworek W.: *System ekspertowy dla identyfikacji modeli ciągłych*. Praca dyplomowa. Wydział Automatyki, Elektroniki i Informatyki. Politechnika Śląska, Gliwice 1988.

52. Juditsky A. et al.: *Nonlinear Black-box Models in System Identification: Mathematical Foundations*. Automatica, Vol. 31, 1725 - 1750, 1995.

53. Kailath T.: *Linear Systems*. Prentice Hall., Englewood Cliffs, New Jersey 1980.

54. Kalman R. E.: *A new approach to linear filtering and prediction problems*. J. Basic Eng., Transactions ASME, Part D, Vol. 82, 35 – 45, 1960.

55. Kanathani K.: *Geometric Information Criterion for Model Selection*. Int. Journal of Computer Vision, vol. 26, 171 – 189, 1998.

56. Kashyap R. L., Rao A. R.: *Dynamic Stochastic Models from Empirical Data*. Academic Press, New York 1976.

57. Kasprzyk J.: *Uniwersalność modeli niskich rzędów dla obiektów ciągłych sterowanych cyfrowo*, Zeszyty Naukowe Pol. Śląskiej, Automatyka 93, 139 - 150, 1989.

58. Kasprzyk J.: *Model Structure Determination in Parametric Model Identification.* Systems Science, vol. 23, no 2, 89 - 95, 1997.

59. Kasprzyk J.: *Modified least squares for output error model identification*, Proceedings of the 5th Int. Symp. on Methods and Models in Automation and Robotics MMAR'98, Międzyzdroje, Vol. 2, 589 – 592, 1998.

60. Kasprzyk J.: *Identyfikacja modeli matematycznych dla celów aktywnego tłumienia drgań.* Materiały IV Szkoły Metody Aktywne Redukcji Drgań i Hałasu, Kraków-Krynica, 89 – 94, 1999.

61. Kasprzyk J.: *Computer Aided Process Identification Using MULTI-EDIP.* Proceedings of the IASTED International Conference Applied Simulation and Modelling, Marbella, 13 – 18, 2001.

62. Kasprzyk J. (red.): *Identyfikacja procesów*, Wydawnictwa Politechniki Śląskiej, Gliwice 2002.

63. Kasprzyk J.: *Model Identification for Active Noise Control System Design.* Proceedings of the 6th Conference on Active Noise and Vibration Control Methods, paper no. 015 (CD), Cracow 2003.

64. Kasprzyk J.: *MULTI-EDIP – An Interactive Software Package For Process Identification.* Proceedings of the 13th IFAC Symposium on System Identification (SYSID 2003), 1484 – 1489, Rotterdam 2003.

65. Kasprzyk J.: *Model identification for Active Noise Control – a case study.* Archives of Control Sciences, Vol. 14, No. 3, 219 – 243, 2004.

66. Kasprzyk J.: *Model Identification for Active Noise Control in the Presence of Primary Noise.* Mechanics, Vol. 24, No 2, 73 – 76, 2005.

67. Kay S. M.: *Modern Spectral Estimation: Theory and Applications.* Prentice Hall, Englewood Cliffs, 1986.

68. Kollar I., Pintelon R., Rolain Y., Schoukens J., Simon G.: *Frequency Domain System Identification Toolbox for Matlab: Automatic Processing – from Data to Models.* Proceedings of the 13th IFAC Symposium on System Identification (SYSID 2003), 1502 – 1506, Rotterdam 2003.

69. Krawczyk J.: *Pakiet MULTI-EDIP w systemie Windows.* Praca dyplomowa. Wydział Automatyki, Elektroniki i Informatyki. Politechnika Śląska, Gliwice 1996.

70. Kristensen N. R., Madsen H., Jørgensen S. B.: *Parameter estimation in stochastic grey-box models.* Automatica, Vol. 40, 225 – 237, 2004.

71. Kullback S., Leibler R. A.: *On information and sufficiency.* Annals of Mathematical Statistics, Vol. 22, 79-86, 1951.

72. Kuo S. M., Morgan D. R.: *Active Noise Control Systems. Algorithms and DSP Implementations.* J. Wiley & Sons, New York 1996.

73. Kurz H., Von: *Erprobung und Vergleich von parameter-adaptiven Regel algorithmen bei verschiedenen Prozessen.* Regelungstechnik, H1, 1980.

74. Larimore W. E.: *ADAPT$_X$ Automated System Identification Software. Users Manual.* Adaptics, Inc., 40 Fairchild Drive, Reading, MA, 2002.

75. Larson J. E., Persson P.: *The Konowledge Database Used in an Expert System Interfacefor Idpac.* Preprints of IFAC Workshop Artificial Intelligence in Real-Time Control, Swansea, 107 – 112, 1988.

76. Lehman E. L.: *Testing Statistical Hypotheses.* Wiley, London 1959.

77. Leontaritis I. J., Billings S. A.: *Model selection and validation methods for non-linear systems.* Int. J. Control, Vol. 41, 303 – 341, 1987.

78. Ljung L.: *Asymptotic behavior of the Extended Kalman Filter as a parameter estimator for linear systems.* IEEE Trans. on Automatic Control, Vol. 24, 36 – 50, 1979.

79. Ljung L.: *System identification: Theory for the User.* (Second ed.), Prentice-Hall, Upper Saddle River, New Jersey 1999.

80. Ljung L.: *Educational Aspects of Identification Software User Interfaces.* Proceedings of 13th IFAC Symposium on System Identification (SYSID 2003), 1590 – 1594, Rotterdam 2003.

81. Ljung L.: *System Identification Toolbox for use with MATLAB. Version 6.* The MathWorks Inc., Natick, MA, 2004.

82. Ljung L., Söderström T.: *Theory and Practice of Recursive Identification.* MIT Press, Cambridge, Mass., 1983.

83. Ljung L., Wahlberg B.: *Asymptotic properties of the least-squares method for estimating transfer functions and disturbance spectra.* Advance Applied Probability, Vol. 24, 412 – 440, 1992.

84. Lueg P.: *Process of silencing sound oscillation.* U.S. Patent No. 2 043 416, 1936.

85. Macchi O.: *Adaptive Processing. The Least Mean Squares Approach with Applications in Transmission.* J. Wiley & Sons, Chichester 1995.

86. Mańczak K., Nahorski Z.: *Komputerowa identyfikacja obiektów dynamicznych.* PWN, Warszawa 1983.

87. Marvasti Farokh A.: *Nonuniform Sampling: Theory and Practice.* Kluwer Academic Plenum Publishers, New York 2001.

88. Mayer zu Farwig H., Unbehauen H.: *Application of Expert System Technique in the Field of System Identification – A Survey.* Preprints of Int. Conference on System Science, Wrocław 1989.

89. McKelvey T.: *Frequency Domain Identification.* Proceedings of the 12th IFAC Symposium on System Identification, Vol. 1., 7 – 18, Santa Barbara, CA, 2000.

90. Michalczyk M.: *Adaptive Control Algorithms for Three-Dimensional Zones of Quiet.* Jacek Skalmierski Computer Studio, Gliwice 2004.

91. Mohler R.: *Nonlinear systems: Applications to bilinear control.* Prentice Hall, 1991.

92. Monsion M. et al.: *An Expert System for Industrial Process Identification.* Preprints of IFAC Workshop Artificial Intelligence in Real-Time Control, Swansea, 95 – 99, 1988.

93. Nahorski Z.: *Identyfikacja obiektów z czasem ciągłym na podstawie zakłóconych dyskretnych pomiarów przebiegów przejściowych.* PWN, Warszawa 1991.

94. Nelles O.: *Nonlinear System Identification. From Classical Approaches to Neural Networks and Fuzzy Models.* Springer Verlag, Berlin-Heidelberg 2001.

95. Nelson P. A., Eliott S. J.: *Active Control of Sound.* Academic Press, London 1992.

96. Nguyen V. V., Wood E. G.: *Review and unification of linear identifiability.* SIAM Rev., vol. 24, 34 – 51, 1982.

97. Niederliński A., Kasprzyk J., Figwer J.: *EFPI - An Integrated Software Environment for System- and Signal Identification.* Proceedings of the 9th IFAC/IFORS Symposium on Identification and System Parameter Estimation, Budapest, 567 - 572, 1991.

98. Niederliński A., Kasprzyk J., Figwer J.: *EDIP - Ekspert dla identyfikacji procesów. Wersja 2. Podręcznik użytkownika.* Wydawnictwo Politechniki Śląskiej, Gliwice 1992.

99. Niederliński A., Kasprzyk J., Figwer J.: *Computer Aided Process Identification.* Computer Aided Control System Design. (M. Brdyś and K. Malinowski, Ed.), 73 - 98. World Scientific Publishing Co. Pte. Ltd., Singapore 1994.

100. Niederliński A. Kasprzyk J., Figwer J.: *Inteligentna identyfikacja dynamicznych modeli matematycznych typu ARX - drzewo decyzyjne lub system ekspertowy?* Problemy sztucznej inteligencji (W. Traczyk, red.), 157 – 172, Wiedza i Życie, Warszawa 1995.

101. Niederliński A., Kasprzyk J., Figwer J.: *MULTI-EDIP - Analizator Wielowymiarowych Sygnałów i Obiektów.* Wydawnictwo Politechniki Śląskiej, Gliwice 1997.

102. Niederliński A., Kasprzyk J., Figwer J.: *A Strategy And Tool For Computer Aided Process Identification*", Automatyka, Sterowanie, Zarządzanie, pod red. J. Gutenbauma, 309 - 324, PAN – IBS, Warszawa 2002.

103. Niederliński A., Kasprzyk J., Pawełczyk M.: *Lambda-Monitoring for RLS Structure Identification of MIMO ARX Plants.* Proceedings of the 2nd Int. Symposium on Methods and Models in Automation and Robotics MMAR'95, Międzyzdroje, Vol. 1, 285 - 290, Wydawnictwo Uczelniane Politechniki Szczecińskiej, Szczecin 1995.

104. Niederliński A., Mościński J.: *The influence of sampling interval on discrete system parameter identifiability.* Archiwum Automatyki i Telemechaniki, Vol. 30, nr 2, 147 – 158, 1985.

105. Niemiec W.: *Badania symulacyjne wpływu warunków prowadzenia eksperymentu na wyniki identyfikacji.* Praca dyplomowa. Wydział Automatyki, Elektroniki i Informatyki. Politechnika Śląska, Gliwice 2004.

106. Nikias C.L., Petropulu A.P.: *Higher-Order Spectra Analysis – A Nonlinear Signal Processing Framework.* PTR Prentice Hall, Englewood Cliffs, New Jersey, 1993.

107. Norton J. P.: *An Introduction to Identification.* Academic Press, London 1986.

108. Norton J. P., Kasprzyk J.: *Process Identification.* Advanced Control with Matlab and Simulink (J. Mościński, Z. Ogonowski, Ed.), 108 - 149, Ellis Horwood, 1995.

109. Osier D., Grobman S., Batson S.: *Teach Yourself Delphi 2 in 21 Days.* SAMS Publishing Co., 1996.

110. Otnes R. K., Enochson L.: *Analiza numeryczna szeregów czasowych.* WNT, Warszawa 1978.

111. Ouderaa E. Van der, Schoukens J. and Renneboog J.: *Peak factor minimization of input and outut signals of linear systems.* IEEE Trans. on Instrum. Meas, Vol. 37, 207 – 212, 1988.

112. Passino K. M.: *Bridging the Gap between Conventional and Intelligent Control.* IEEE Control Systems, Vol. 13, 12 – 18, 1993.

113. Pawełczyk M.: *Active Noise Control for Compact Acoustic Plants.* J. Skalmierski Computer Studio, Gliwice 1999.

114. Peterka V.: *Bayesian system identification.* Automatica, Vol. 17, 41 – 53, 1981.

115. Pintelon R., Schoukens J.: *System Identification. A Frequency Domain Approach.* IEEE Press, New York 2001.

116. Popper K. R.: *The Logic of Scientific Discovery.* Hutchinson & Co. (Publisher) Ltd., London 1980.

117. Pötscher B. M., Srinivasan S.: *A comparison of order estimation procedures for ARMA models.* Statistica Sinica, vol. 4, 29 – 50, 1994.

118. Rissanen J.: *Modelling by Shortest Data Description.* Automatica, Vol. 14, 465 – 471, 1978.

119. Rolain Y., Schoukens J. and Pintelon R.: *Order Estimation for Linear Time-invariant Systems using Frequency Domain Identification Methods.* Proceedings of the 34th Conference on Decision & Control, 3588 - 3593, New Orleans, LA, December 1995.

120. Savage L. J.: *The Foundations of Statistical Inference.* Methuen, London, 1962.

121. Schetzen M.: *The Volterra and Wiener Theories of Nonlinear Systems.* Wiley, New York 1980.

122. Schoukens J., Dobrowiecki T., Pintelon R.: *Parametric and Nonparametric Identification of Linear Systems in the Presence of Nonlinear Distortions – A Frequency Domain Approach.* IEEE Trans. on Aut. Control, Vol. AC-43, 176 – 190, 1998.

123. Schroeder M. R.: *Synthesis of low-peak factor signals and binary sequences with low autocorrelation.* IEEE Trans. Information Theory, Vol. 16, 85 – 89, 1970.

124. Schuster H. G.: *Deterministic Chaos. An Introduction.* VCH Verlagsgesellschaft, 1988.

125. Schwarz G.: *Estimating the dimension of a model.* The Annals of Statistics, Vol. 6, 461 – 464, 1978.

126. Shannon C. E.: *A Mathematical Theory of Communication.* Bell System Technical Journal, Vol. 27, 379 – 423, 623 – 656, 1948.

127. Simon G., Schoukens J., Rolain Y.: *Automatic model selection for linear time-invariant systems.* Proceedings of the 12th IFAC Symposium on System Identification (SYSID 2000), vol. 1, 379 – 384, Santa Barbara, CA, 2000.

128. Sjöberg J. et al.: *Nonlinear Black-box Modeling in System Identification: a Unified Overwiew.* Automatica, Vol. 31, 1691 - 1724, 1995.

129. Söderström T.: *On model structure testing in system identification.* Int. J. Control, Vol. 26, 1 – 18, 1977.

130. Söderström T.: *On a method for model structure selection in system identification.* Automatica, Vol. 17, 387 – 388, 1981.

131. Söderström T., Stoica P.: *Instrumental Variable Methods for Systems identification.* Lecture Notes in Control and Information Sciences, Springer-Verlag, New York 1983.

132. Söderström T., Stoica P.: *System Identification.* Prentice Hall International, London 1989.

133. Söderström T., Stoica P.: *On Covariance Function Tests Used in System Identification.* Automatica, Vol. 26, 125 – 133, 1990.

134. Sugiyama M.: *Functional Analytic Framework for Model Selection.* Proceedings of the 13th IFAC Symposium on System Identification (SYSID 2003), 73 – 78, Rotterdam 2003.

135. Texeira S., Pacheco R.: *DELPHI 4. Vademecum Profesjonalisty.* Helion, Gliwice 1999.

136. Therrien Ch.: *Discrete Random Signals and Statistical Signal Processing*, Prentice Hall Int. Inc., New Jersey 1992.

137. Tong H.: *Non-linear time series.* Clarendon Press, Oxford 1993.

138. Topor Z., Johannsen L., Kasprzyk J., Remmers J.: *Dynamic Ventilatory Response to CO_2 in Congestive Heart Failure Patients with and without Central Sleep Apnea.* Journal of Applied Physiology, No 91, 408 - 416, 2001.

139. Tugnait J. K., Tontiruttananon C.: *Identification of linear system via spectral analysis given time-domain data: Consistency, reduced-order approximation and performance analysis.* IEEE Trans. on Aut. Control, Vol. AC-43, 1354 – 1373, 1998.

140. Tulleken H. A. J. F.: *Grey-box modelling and identification using physical knowledge and Bayesian techniques.* Automatica, Vol. 29, 285 – 308, 1993.

141. Van Overschee P., De Moor B., Aling H., Kosut R., Boyd S.: *Xmath Interactive Identification Module.* Integrated Systems, Inc., Santa Clara, CA, 1994.

142. Wilks S. S.: *Mathematical Statistics*. Wiley, London 1962.

143. Zhu Y.: *Use of Error Criteria in Identification for Control*. Proceedings of the 12th IFAC Symposium on System Identification (SYSID 2000), vol. 1, 307 – 312, Santa Barbara, CA, 2000.

Dodatek A

Estymacja parametrów modeli liniowych

W praktyce najczęściej wyznaczenie modelu M sprowadza się do estymacji wektora parametrów θ dla założonej struktury $\mathcal{M}$. Dla stochastycznych modeli liniowych odbywa się to na ogół na drodze minimalizacji funkcji strat definiowanej jako suma kwadratów błędów predykcji:

$$\hat{\theta} = \arg\min_{\theta} \frac{1}{N} \sum_{i} \varepsilon^2 (i \mid i-1), \tag{A.1}$$

gdzie dla modelu *ogólnego MISO* błąd predykcji można wyznaczyć z równania:

$$\varepsilon(i \mid i-1) = \frac{D(z^{-1})}{C(z^{-1})} \left[A(z^{-1}) y(i) - \sum_{k=1}^{q} \frac{z^{-d_k} B_k(z^{-1})}{F_k(z^{-1})} u_k(i) \right]. \tag{A.2}$$

Wprowadzając oznaczenia:

$$w_k(i) = z^{-d_k} \frac{B_k(z^{-1})}{F_k(z^{-1})} u_k(i) \;,\; k = 1, 2, \ldots, q \tag{A.3}$$

oraz

$$v(i) = A(z^{-1}) y(i) - \sum_{k=1}^{q} w_k(i) \tag{A.4}$$

równanie (A.2) najwygodniej jest przedstawić w postaci formy liniowej

$$\varepsilon(i \mid i-1) = y(i) - \varphi^T (i-1)\theta, \tag{A.5}$$

gdzie wektor danych

$$\begin{aligned} \varphi^T(i-1) = [& -y(i-1) \;\; -y(i-2) \;\; \ldots \;\; -y(i-dA) \\ & u_1(i-d_1) \;\; u_1(i-d_1-1) \;\; \ldots \;\; u_1(i-d_1-dB_1) \\ & u_2(i-d_2) \;\; u_2(i-d_2-1) \;\; \ldots \;\; u_2(i-d_2-dB_2) \\ & \vdots \\ & u_q(i-d_q) \;\; u_q(i-d_q-1) \;\; \ldots \;\; u_q(i-d_q-dB_q) \end{aligned}$$

$$
\begin{array}{l}
w_1(i-1)\ \ w_1(i-2)\ \ \ldots\ \ w_1(i-dF_1) \\
w_2(i-1)\ \ w_2(i-2)\ \ \ldots\ \ w_2(i-dF_2) \\
\vdots \\
w_q(i-1)\ \ w_q(i-2)\ \ \ldots\ \ w_q(i-dF_q) \\
\varepsilon(i-1\,|\,i-2)\ \ \varepsilon(i-2\,|\,i-3)\ \ \ldots\ \ \varepsilon(i-dC\,|\,i-dC-1) \\
-v(i-1)\ \ -v(i-2)\ \ \ldots\ \ -v(i-dD)\ ]
\end{array}
\qquad \text{(A.6)}
$$

a wektor parametrów

$$
\begin{array}{l}
\theta^T = [\ a_1\ \ a_2\ \ \ldots\ \ a_{dA} \\
b_{0,1}\ \ b_{1,1}\ \ \ldots\ \ b_{dB_1,1} \\
b_{0,2}\ \ b_{1,2}\ \ \ldots\ \ b_{dB_2,2} \\
\vdots \\
b_{0,q}\ \ b_{1,q}\ \ \ldots\ \ b_{dB_q,q} \\
f_{1,1}\ \ f_{2,1}\ \ \ldots\ \ f_{dF_1,1} \\
f_{1,2}\ \ f_{2,2}\ \ \ldots\ \ f_{dF_2,2} \\
\vdots \\
f_{1,q}\ \ f_{2,q}\ \ \ldots\ \ f_{dF_q,q} \\
c_1\ \ c_2\ \ \ldots\ \ c_{dC} \\
d_1\ \ d_2\ \ \ldots\ \ d_{dD}\ ].
\end{array}
\qquad \text{(A.7)}
$$

Dla modeli *ARX*, *FIR* oraz *AR* błąd predykcji $\varepsilon(i\,|\,i\text{-}1)$ jest liniową funkcją wszystkich parametrów, co umożliwia estymację wektora $\hat{\theta}$ za pomocą bezpośredniej metody najmniejszych kwadratów *LS* (*Least Squares*), por. A.1.

Dla modeli *ARMAX*, *OE*, *BJ*, *ogólnego*, *MA* i *ARMA* występuje nieliniowa zależność błędu predykcji względem niektórych elementów wektora θ, stąd do rozwiązania zadania (A.1) zaproponowano wiele różnych algorytmów, których przegląd można znaleźć np. w [79]. Większość z nich jednak da się sprowadzić do rekurencyjnej metody błędu predykcji *RPE* (*Recursive Prediction Error*), por. A.1.

Jeżeli natomiast wybrany został model transmitancyjny toru sterowania, to w systemie *MULTI-EDIP* zakłada się milcząco, że wtedy celem identyfikacji nie jest minimalizacja błędu predykcji, ale uzyskanie ocen asymptotycznie zbieżnych do rzeczywistych parametrów toru sterowania. Stąd w przypadku identyfikacji obiektu *SISO* parametry transmitancji toru sterowania $G\left(z^{-1}\right)= z^{-d}\,\dfrac{B\left(z^{-1}\right)}{F\left(z^{-1}\right)}$ wyznacza się korzystając z metody zmiennej pomocniczej (instrumentalnej) *IV* (*Instrumental Variable*), por. [101, 131]. Jeżeli identyfikowany jest obiekt *MISO*, to modele transmitancji torów sterowania identyfikowane są metodą polegającą na iteracyjnej estymacji parametrów poszczególnych torów metodą *IV*, przy czym za każdym razem oblicza się ocenę wyjścia k–tego toru jako różnicę między wyjściem $y(i)$ a wyjściami pozostałych torów obliczonymi dla poprzedniej iteracji, por. [24, 101].

A.1. Implementacja metody najmniejszych kwadratów

W przypadku liniowej zależności błędu predykcji od parametrów θ rozwiązanie zadania (A.1) odbywa się bezpośrednią metodą najmniejszych kwadratów, polegającą na rozwiązaniu układu równań normalnych Gaussa:

$$
\left[\frac{1}{N}\sum_{i=i_0}^{N}\varphi(i-1)\varphi^T(i-1)\right]\hat{\theta}=\frac{1}{N}\sum_{i=i_0}^{N}\varphi(i-1)y(i), \qquad \text{(A.8)}
$$

gdzie wektor φ zawiera, zgodnie z (A.6), elementy odpowiednie do wymiarowości wektora wejść oraz typu modelu (*FIR*, *ARX* lub *AR*, por. 4.2.1) i jego struktury, a dolny zakres sumowania i_0 jest tak dobrany, aby indeksy elementów wektora $\varphi(i-1)$ były dodatnie, tj. $i_0 = 1 + \max(dA, d_1 + dB_1, d_2 + dB_2, \ldots, dq + dBq)$.

Jednak wyznaczanie tzw. macierzy momentów $\Phi^T\Phi=\dfrac{1}{N}\sum_{i=i_0}^{N}\varphi(i-1)\varphi^T(i-1)$ i wektora $\Phi^T Y=\dfrac{1}{N}\sum_{i=i_0}^{N}\varphi(i-1)y(i)$ bezpośrednio z przedstawionych wzorów byłoby numerycznie nieefektywne. Stąd w systemie *MULTI-EDIP* zaproponowano specyficzny sposób ich obliczania, wykorzystując fakt, że zawierają one wartości funkcji autokorelacji i korelacji wzajemnych wejść i wyjść obliczane dla odpowiednio przesuniętych w czasie danych.

Dlatego też najpierw wyznacza się wszystkie potrzebne korelacje $\rho_{yy}(0)$ do $\rho_{yy}(dA)$, $\rho_{u_k u_k}(0)$ do $\rho_{u_k u_k}(dB_k)$, $\rho_{yu_k}(d_k - dA)$ do $\rho_{yu_k}(d_k + dB_k)$, oraz $\rho_{u_j u_k}(d_k - d_j - dB_j)$ do $\rho_{u_j u_k}(d_k + dB_k - d_j)$ dla $k, j = 1, 2, \ldots, q$, gdzie poszczególne korelacje oblicza się zgodnie z zależnością

$$\rho_{xy}(l)=\begin{cases}\dfrac{1}{N-l}\displaystyle\sum_{i=l+1}^{N}x(i)y(i-l), & l\geq 0\\[2ex] \dfrac{1}{N+l}\displaystyle\sum_{i=1}^{N+l}x(i)y(i-l), & l<0\end{cases} \tag{A.9}$$

Następnie tworzy się macierz $\Phi T\Phi$ i wektor ΦTY wstawiając odpowiednie wartości korelacji skorygowane ze względu na granice sumowania wynikające z (A.8) i (A.6). Przykładowo, element (1,1) macierzy $\Phi T\Phi$ powinien zawierać $\rho_{yy}(0)$ obliczone dla sumy od i_0 do $N-1$, element (2,2) natomiast $\rho_{yy}(0)$ obliczone od i_0-1 do $N-2$, itd.

Zastosowany tu sposób wyznaczania macierzy $\Phi^T\Phi$ i wektora Φ^TY ma szczególne znaczenie w przypadku identyfikacji modelu o nieznanej strukturze, gdyż pozwala niezwykle efektywnie rozwiązać zadanie estymacji parametrów modeli dla kolejnych testowanych struktur. Aby wyznaczyć macierz momentów zgodnie z zaproponowanym algorytmem, trzeba wykonać około m razy mnożeń mniej dla jednej struktury, gdzie m jest liczbą estymowanych parametrów. Jednak w przypadku estymacji parametrów dla wielu różnych struktur macierze korelacji wyznacza się tylko raz na początku, aczkolwiek dla zakresu przesunięć l wynikającego z całego przewidywanego zakresu zmian stopni wielomianów $n_{\max}$ i opóźnień $d_{\max}$. Jednak później wyznaczenie macierzy $\Phi^T\Phi$ i wektora Φ^TY dla każdej badanej struktury opiera się na już obliczonych wartościach korelacji, stąd przyspieszenie obliczeń może wynieść tu od dziesiątek do tysięcy razy, w zależności od wymiarowości problemu. Ponadto obliczone korelacje mogą być później wykorzystane jeszcze do badania uwarunkowania numerycznego zadania estymacji (A.1), por. dodatek B.

Oprócz numerycznej efektywności obliczania macierzy w (A.8) istotne jest także numerycznie poprawne rozwiązanie tego układu równań.

A.2. Implementacja rekurencyjnej metody błędu predykcji

Jeżeli w (A.2) występuje nieliniowa zależność błędu predykcji względem niektórych elementów wektora θ, to do rozwiązania zadania (A.1) można zastosować metodę Gaussa-Newtona, co prowadzi do algorytmu *RPE* (*Recursive Prediction Error*). Wyprowadzony w [81] algorytm *RPE* dla obiektów *SISO* został w systemie *MULTI-EDIP* rozszerzony na obiekty *MISO*, por. [101]. Jednocześnie w celu zmniejszenia wpływu warunków początkowych na wynik końcowy wprowadzono modyfikację polegającą na iteracyjnym powtarzaniu algorytmu dla tych samych danych, przy czym po każdym rekurencyjnym przetworzeniu wszystkich pomiarów uzyskana ocena $\hat{\theta}$ jest traktowana jako warunek początkowy dla ponownego przetwarzania

pomiarów za pomocą *RPE*. W ten sposób po kilku takich iteracjach uzyskuje się zazwyczaj oceny, które praktycznie już nie zmieniają się dla kolejnych iteracji. Maksymalną liczbę iteracji ustalono arbitralnie na 20.

Ponadto zaproponowany w [81] algorytm badania stabilności wielomianu $C(z^{-1})$, który w przypadku wykrycia niestabilności umożliwia taką modyfikację jego współczynników, aby „ściągnąć" pierwiastki wielomianu do wnętrza koła jednostkowego, został zaimplementowany także do wielomianów $D(z^{-1})$ i $Fk(z^{-1})$, gdyż jak wykazały badania, w przypadku wystąpienia w widmie sygnału wyjściowego kilku pików rezonansowych wielomiany te bardzo łatwo tracą stabilność, co uniemożliwia poprawną filtrację zgodnie z (A.3), a także filtrację odpowiednich sygnałów w celu obliczenia ich gradientów.

Wprawdzie algorytm *RPE* jest tu stosowany do estymacji parametrów modelu stacjonarnego, jednak ze względu na własności numeryczne związane z filtracją przez wielomiany, które na początku rekurencji są mało dokładne, zastosowano tu współczynnik zapominania λ mniejszy od jedynki. W przypadku identyfikacji obiektu współczynnik ten zmienia się w czasie zgodnie z zależnością

$$\lambda(i)=0.99\lambda(i-1)+0.01,\quad \text{dla } \lambda(0)=0.95, \tag{A.10}$$

natomiast dla ciągu czasowego

$$\lambda(i)=0.98\lambda(i-1)+0.02,\quad \text{dla } \lambda(0)=0.5. \tag{A.11}$$

Wartości początkowe oraz współczynniki filtracji zostały dobrane w sposób doświadczalny na drodze badań symulacyjnych. Mniejsze wartości współczynnika λ dla ciągu czasowego wynikają stąd, że występuje tu gorsza dokładność uzyskiwanych ocen parametrów na początku, ponieważ estymacja odbywa się tylko na podstawie znajomości sygnału wyjściowego procesu. W przypadku obiektu wykorzystanie w algorytmie wartości wejść umożliwia szybsze uzyskanie ocen zbliżonych do poprawnych.

Podobnie, ponieważ na początku rekurencji wielomian $C(z^{-1})$ jest wyznaczony mało dokładnie, podczas gdy jego wpływ na filtrację odpowiednich sygnałów jest bardzo istotny, stąd w [81] zaproponowano wprowadzenie tzw. *współczynnika kontrakcji K* (ang. *contraction factor*). W ten sposób zamiast typowej dla *RPE* filtracji przez wielomian $1/C(z^{-1})$ jest stosowana filtracja $1/C(Kz^{-1})$, a więc dla małej liczby pomiarów algorytm ten staje się czymś pośrednim między metodą *RPE* a metodą regresji pseudoliniowej *PLR*.. W przypadku obiektu współczynnik ten oblicza się według wzoru

$$K(i)=0.98K(i-1)+0.02,\quad \text{dla } K(0)=0, \tag{A.12}$$

a dla ciągu czasowego

$$K(i) = 0.95\,K(i-1) + 0.05, \quad \text{dla } K(0) = 0.5. \tag{A.13}$$

Zależności te dobrano także na drodze doświadczalnej. Tu większa wartość współczynnika K dla ciągu czasowego wynika stąd, że wielomian $C(z^{-1})$ w modelowaniu własności dynamicznych ma w tym przypadku znacznie większe znaczenie niż dla obiektu, gdzie służy on tylko do modelowania toru zakłócenia.

W miarę przebiegu rekurencji oba współczynniki dążą do jedności. Dla większej liczby pomiarów N (np. $N > 500$) już w drugiej iteracji ich wpływ na wynik algorytmu staje się pomijalny.

Dodatek B

Testy uwarunkowania macierzy korelacji

Badanie stopnia uwarunkowania odpowiednio skonstruowanych macierzy korelacji wejść i wyjść może pomóc w ustaleniu przyczyn słabego uwarunkowania równań estymatora. Przedstawione poniżej macierze korelacji zostały skonstruowane dla modelu *MISO ogólnego* opisanego równaniami (4.9), (4.20a) i (4.20b). Tworzenie tych macierzy dla modeli prostszych polega na usunięciu tych elementów z macierzy, które odpowiadają wielomianom nie występującym w danym typie modelu. W systemie *MULTI-EDIP* stopień uwarunkowania tych macierzy definiowany jest jako stosunek wartości osobliwych maksymalnej do minimalnej, a za granicę dobrego uwarunkowania przyjęto wartość 10^4.

Badanie stopnia uwarunkowania *macierzy autokorelacji wyjść*

$$\rho_{yy} = \begin{bmatrix} \rho_{yy}(0) & \rho_{yy}(1) & \cdots & \rho_{yy}(dA + dF_{\max} - 1) \\ \rho_{yy}(1) & \rho_{yy}(0) & \cdots & \rho_{yy}(dA + dF_{\max} - 2) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{yy}(dA + dF_{\max} - 1) & \rho_{yy}(dA + dF_{\max} - 2) & \cdots & \rho_{yy}(0) \end{bmatrix} \tag{B.1}$$

o wymiarach $dA + dF_{\max}$, gdzie $dF_{\max}$ jest maksymalnym stopniem wielomianów $Fk(z^{-1})$, pozwala na wykrycie zbyt krótkiego okresu próbkowania, zwłaszcza gdy wartość funkcji autokorelacji $\rho_{yy}(1)$ jest bliska wartości $\rho_{yy}(0)$. Słabe uwarunkowanie ρ_{yy} może świadczyć także o istnieniu *offsetu* w danych pomiarowych (gdy wartość średnia wyjścia jest większa od jego odchylenia standardowego) albo o zbyt wysokim stopniu wielomianów $A(z^{-1})$ lub $F(z^{-1})$. Test ten jest stosowany zarówno w przypadku identyfikacji obiektów, jak i sygnałów.

Natomiast dla obiektów badanie uwarunkowania odpowiednich macierzy autokorelacji wejść służy sprawdzeniu, czy ciągi wejściowe spełniają wymagania odpowiedniego pobudzenia. Przede wszystkim poszczególne sygnały wejściowe powinny być ciągami nieustannie pobudzającymi stopnia $dA + dFk + dBk + 1$, stąd badane są stopnie uwarunkowania macierzy

$$\rho_{u_k u_k} = \begin{bmatrix} \rho_{u_k u_k}(0) & \rho_{u_k u_k}(1) & \cdots & \rho_{u_k u_k}(dA + dF_k + dB_k) \\ \rho_{u_k u_k}(1) & \rho_{u_k u_k}(0) & \cdots & \rho_{u_k u_k}(dA + dF_k + dB_k - 1) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{u_k u_k}(dA + dF_k + dB_k) & \rho_{u_k u_k}(dA + dF_k + dB_k - 1) & \cdots & \rho_{u_k u_k}(0) \end{bmatrix} \tag{B.2}$$

dla $k = 1, 2, \ldots, q$.

Jednak w przypadku obiektu *MISO* ciągi wejściowe powinny być także wzajemnie nieskorelowane, gdyż w przeciwnym przypadku prawidłowa identyfikacja poszczególnych torów sterowania byłaby niemożliwa. Stąd bardzo ważne jest badanie stopnia uwarunkowania macierzy korelacji własnych i wzajemnych dla wszystkich wejść:

$$\rho_{uu} = \begin{bmatrix} \rho_{u_1u_1} & \rho_{u_1u_2} & \cdots & \rho_{u_1u_q} \\ \rho_{u_2u1} & \rho_{u_2u_2} & \cdots & \rho_{u_2u_q} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{u_qu_1} & \rho_{u_qu_2} & \cdots & \rho_{u_qu_q} \end{bmatrix}, \tag{B.3}$$

gdzie $\rho_{u_ku_k}$ jest macierzą autokorelacji dla k-tego wejścia o wymiarach $(dB_k + 1)\times(dB_k + 1)$ wyznaczaną zgodnie z (B.2), natomiast

$$\rho_{u_ju_k} = \begin{bmatrix} \rho_{u_ju_k}(d_k - d_j) & \rho_{u_ju_k}(d_k + 1 - d_j) & \cdots & \rho_{u_ju_k}(d_k + dB_k - d_j) \\ \rho_{u_ju_k}(d_k - d_j - 1) & \rho_{u_ju_k}(d_k + 1 - d_j - 1) & \cdots & \rho_{u_ju_k}(d_k + dB_k - d_j - 1) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{u_ju_k}(d_k - d_j - dB_j) & \rho_{u_ju_k}(d_k + 1 - d_j - dB_j) & \cdots & \rho_{u_ju_k}(d_k + dB_k - d_j - dB_j) \end{bmatrix}, \tag{B.4}$$

gdzie $\rho_{u_ju_k}(l)$ oblicza się zgodnie z (A.9).

Jeżeli okaże się, że macierz ρuu jest słabo uwarunkowana, a liczba wejść $q > 2$, to badane są stopnie uwarunkowania macierzy korelacji (B.3) obliczane tylko dla poszczególnych par ciągów wejściowych w celu określenia, która para stanowi przyczynę słabego uwarunkowania.

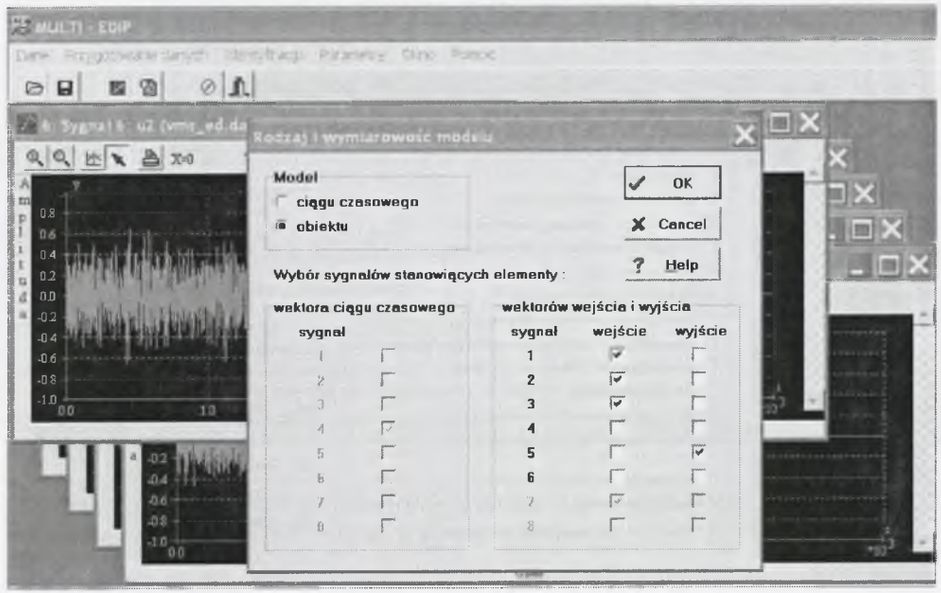
Dodatek C

Praca z systemem MULTI-EDIP

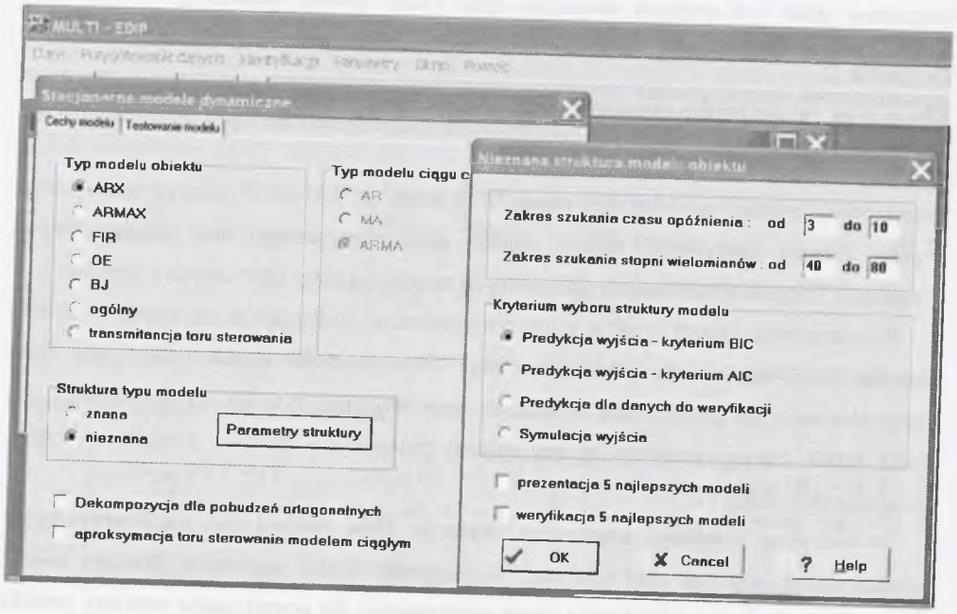
Prezentowany poniżej przykład korzystania z systemu *MULTI-EDIP* dotyczy omówionego w pracy zadania identyfikacji modelu obiektu elektroakustycznego oraz przedstawionych w rozdziale 7 wyników uzyskanych dla zebranych za pomocą karty DSP danych z obiektu.

Po odczytaniu danych z pliku i ich przygotowaniu, polegającym na usunięciu z nich wartości stałej, użytkownik deklaruje rodzaj i wymiarowość modelu określając, które z sygnałów mają być przetwarzane w dalszym ciągu sesji, por. rys. C.1. Tu deklarowany jest obiekt *MISO*, którego wejściem są trzy sygnały pobudzające głośniki, a wyjściem sygnał z drugiego mikrofonu.

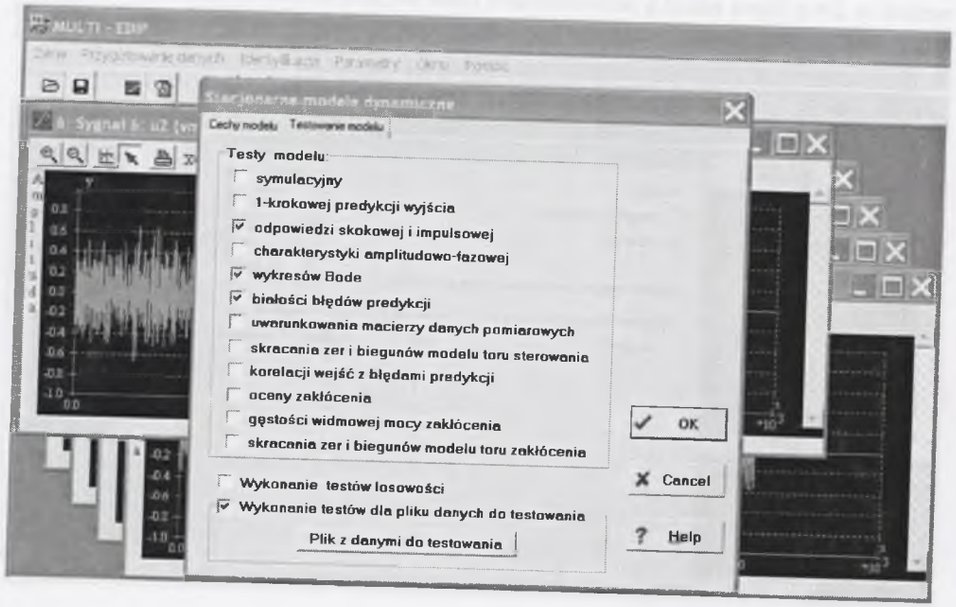
W następnej kolejności użytkownik deklaruje klasę modelu (tu: stacjonarny model parametryczny), jego typ oraz strukturę. W przypadku braku znajomości struktury należy określić zakresy czasów opóźnienia i stopni wielomianów dla poszukiwania struktury modelu oraz kryterium wyboru modelu, por. rys. C.2. Użytkownik wybiera także, które z testów mają być wykonane w celu walidacji i falsyfikacji modelu, por. rys. C.3.



Rys. C.1. Deklaracja sygnałów wejściowych i wyjściowych modelu obiektu
Fig. C.1. Declaration of input and output signals for the model



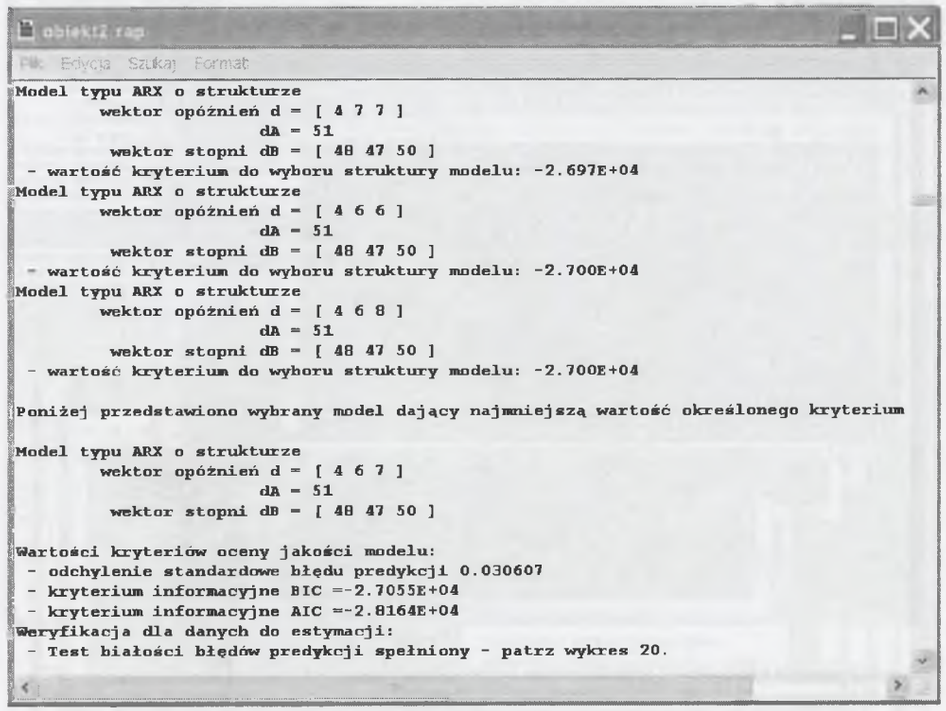
Rys. C.2. Deklaracja typu modelu obiektu, zakresów dla poszukiwania struktury oraz kryterium wyboru modelu
Fig. C.2. Declaration of model type, ranges for structure search and model selection criterion



Rys. C.3. Deklaracja testów dla walidacji i falsyfikacji modelu oraz analizy wyników
Fig. C.3. Declaration of tests for model validation and falsification and for analysis of the results

Po wywołaniu procedury identyfikacji modelu parametrycznego w oknie raportu pojawiają się informacje o kolejnych testowanych strukturach, zgodnie z algorytmem poszukiwań przedstawionym w punkcie 5.3.2, a po zakończeniu procedury informacja o wybranym modelu, por. rys. C.4. W oknie głównym natomiast pojawiają się wykresy przedstawiające wyniki w postaci graficznej, które umożliwiają analizę i dalszą walidację modelu, por. rys. C.5.

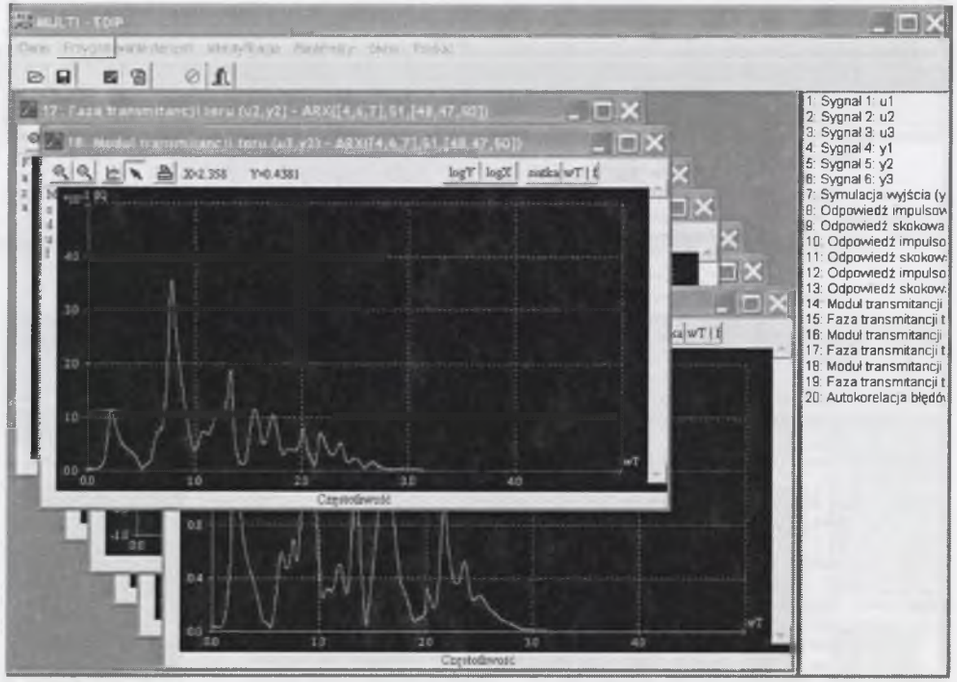
Ponieważ w omawianym przykładzie przyjęto, że w ramach walidacji modelu należy porównać charakterystyki częstotliwościowe otrzymanego modelu z charakterystykami uzyskanymi metodą nieparametryczną dla danych pochodzących z doświadczenia specjalnie przeprowadzonego w tym celu, stąd po odczytaniu danych pochodzących z tego doświadczenia należy określić, jakie modele nieparametryczne mają być identyfikowane, por. rys. C.6, a także inne opcje, np. sposób okienkowania, prezentacji wyników itp.



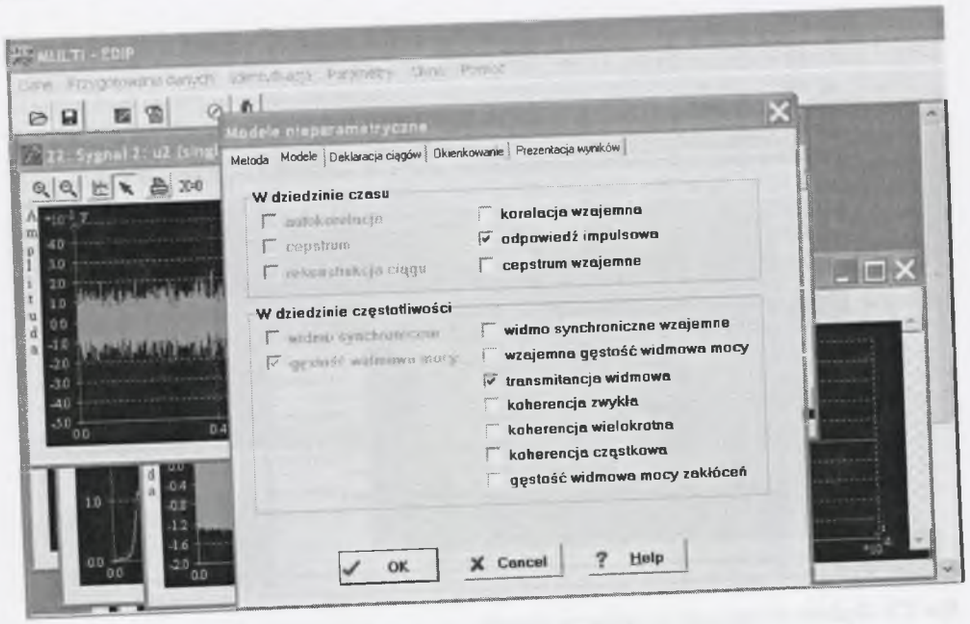
Rys. C.4. Fragment okna raportu z informacją o testowanych strukturach i wybranym modelu
Fig. C.4. A part of the Report Window with information about tested structures and selected model

Po uzyskaniu nowych wyników użytkownik może, korzystając z menu obsługi wykresów, porównywać odpowiednie charakterystyki kopiując wykresy z jednego okna do drugiego, por. rys. C.7 i C.8. W ten sposób było możliwe np. uzyskanie wykresów przedstawionych na rysunkach 7.1 czy 7.2. Natomiast korzystając z opcji odejmowania wyników ze schowka użytkownik może także otrzymać wykres pokazujący różnicę innych wykresów, por. rys. C.9, co np. pozwoliło na obliczenie wskaźników jakości modeli przedstawionych w tabeli 7.1.

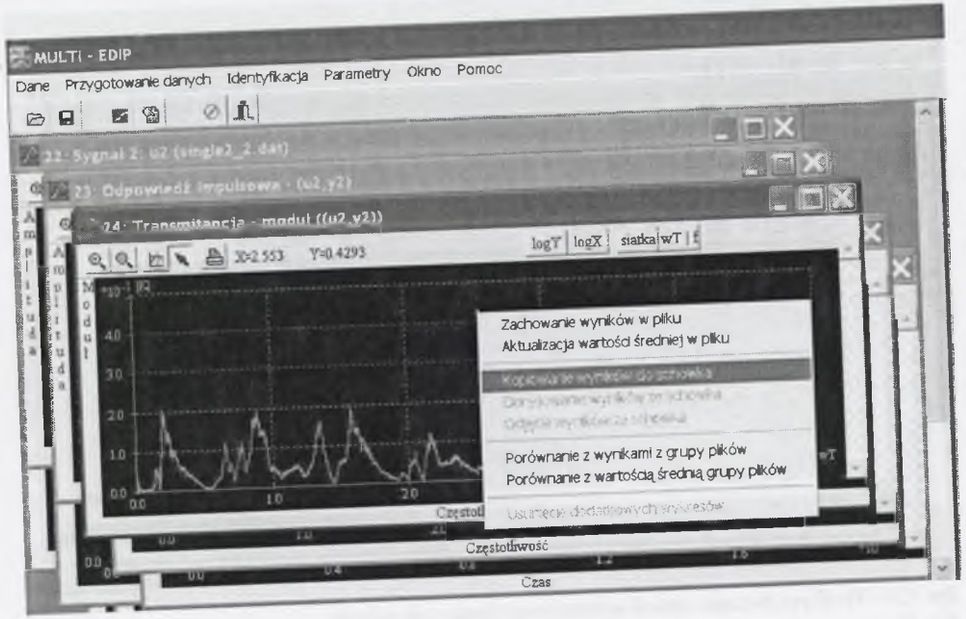
Z kolei na rys. C.10 pokazano sposób deklaracji parametrów dla wstępnego przetwarzania danych umożliwiającego zastosowanie metody *DAF* w przypadku identyfikacji w obecności zakłóceń o widmie wąskopasmowym, por. 7.2. Częstotliwości filtru wyznaczono na podstawie analizy widmowej przeprowadzonej dla sygnału wyjściowego.



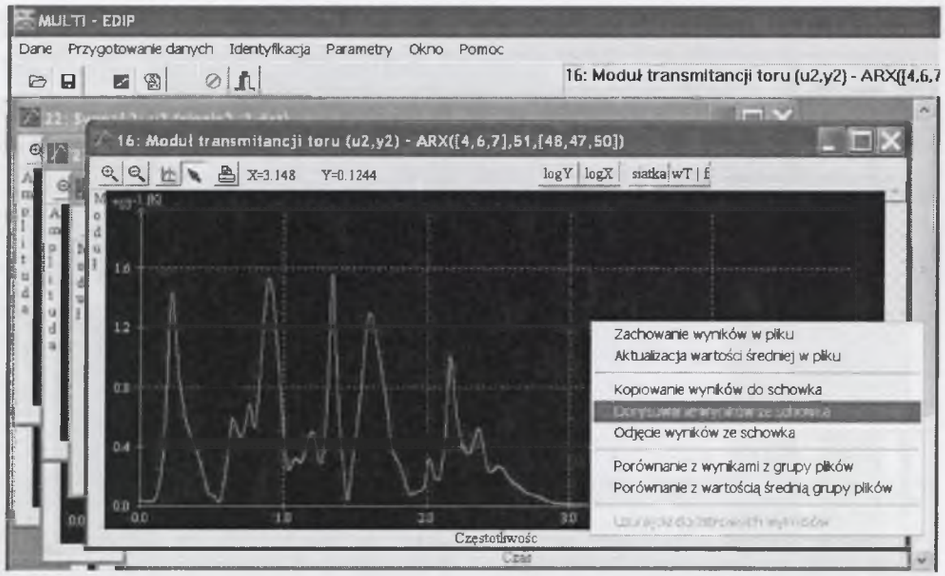
Rys. C.5. Wyniki identyfikacji przedstawione w formie graficznej
Fig. C.5. Results of identification presented in graphical form



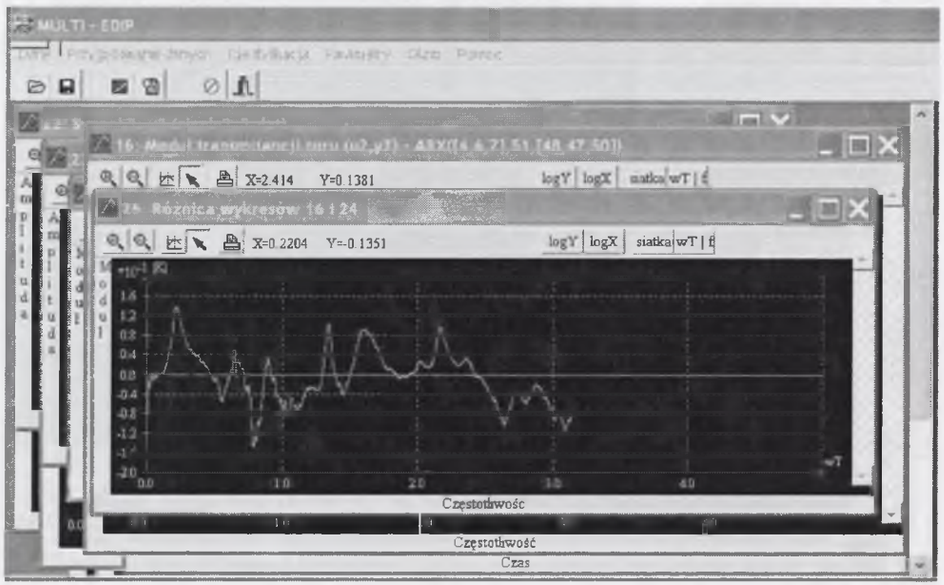
Rys. C.6. Deklaracja modeli nieparametrycznych
Fig. C.6. Declaration of non-parametric models



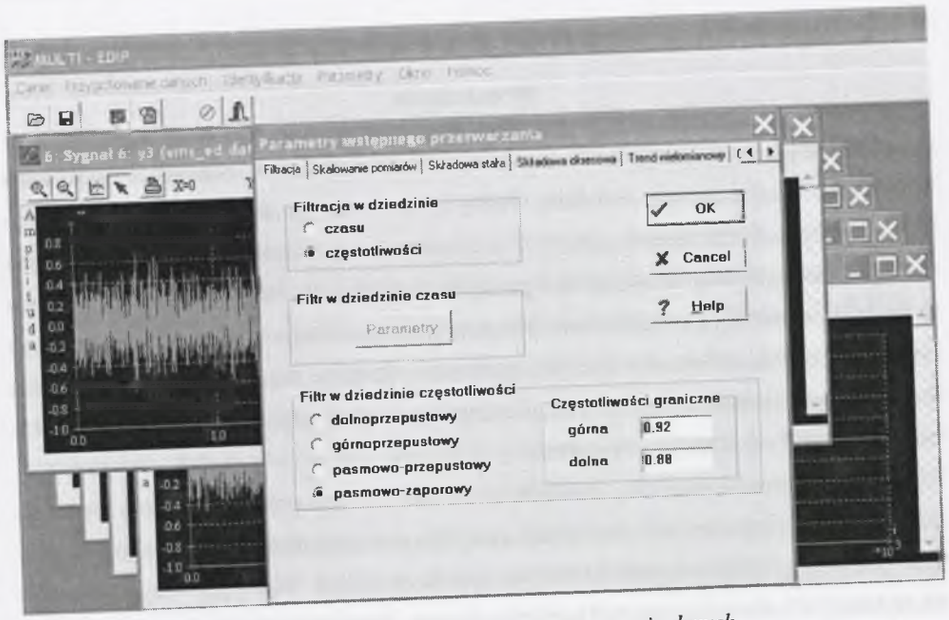
Rys. C.7. Menu obsługi wykresu – żądanie kopiowania do schowka
Fig. C.7. Menu of diagram service – 'copy a diagram to clipboard' request



Rys. C.8. Żądanie dorysowania wyników ze schowka
Fig. C.8. 'Add a diagram from clipboard' request



Rys. C.9. Wynik porównania dwóch wykresów na drodze ich odjęcia
Fig. C.9. Comparison of two diagrams via subtraction



Rys. C.10. Deklaracja parametrów filtru dla wstępnego przetwarzania danych
Fig. C.10. Declaration of filter parameters for data preprocessing

Komputerowe wspomaganie identyfikacji procesów

Streszczenie

Identyfikacja procesów polega na tworzeniu *modeli matematycznych* opisujących właściwości tych procesów na podstawie posiadanej wiedzy *a priori* oraz pomiarów sygnałów zebranych w trakcie *doświadczeń identyfikacyjnych*. Praktyczne stosowanie identyfikacji wymaga dostępu do odpowiedniego oprogramowania oraz pewnych umiejętności wynikających z doświadczenia i wiedzy na temat zasad wnioskowania. Tematem pracy jest analiza problemu komputerowego wspomagania podczas wykonywania procedury identyfikacji, w szczególności dyskusja zadań i praktycznych możliwości implementacji takiego systemu, w tym także ograniczeń, na jakie można napotkać w trakcie jego stosowania.

Na ogół można oddzielić *etap identyfikacji procesu* od *etapu stosowania modelu*. Na etapie identyfikacji wykonywany jest *eksperyment identyfikacyjny*, podczas którego obiekt poddawany jest pobudzeniu, jeśli to możliwe, to w sposób specjalnie zaprojektowany dla celów identyfikacji, a zachowanie procesu jest obserwowane. Etap ten kończy się skonstruowaniem modelu matematycznego, który w zwięzły sposób opisuje zachowanie procesu w czasie eksperymentu. Opis taki powinien być *reprezentatywny*, tj. obowiązywać także na etapie stosowania i być odpowiedni do postawionego celu. W procedurze identyfikacji można wyróżnić następujące etapy: wykonanie eksperymentu, przyjęcie struktury modelu, estymacja modelu zgodnie z wybranym kryterium aproksymacji, testowanie otrzymanego modelu. Jeżeli wynik identyfikacji nie spełnia wymagań, to należy poprawić strukturę modelu i powtórzyć procedurę od kroku drugiego, aż do uzyskania odpowiednich wyników.

Istotę identyfikacji można przedstawić w ujęciu *probabilistycznym*, opartym na wzorze Bayesa. Podejście takie pozwala na przedstawienie spójnej teorii identyfikacji procesów opartej na wnioskowaniu statystycznym, ale ze względu na trudności obliczeniowe ma głównie znaczenie *teoretyczne*. Ponieważ w pracy główny nacisk został położony na *praktyczne* aspekty identyfikacji, stąd wykorzystuje się tu raczej opis w postaci modelu *stochastycznego*, w którym niepewność jest przedstawiana za pomocą tzw. innowacji, będących elementarnymi zmiennymi losowymi o zadanym rozkładzie prawdopodobieństwa. W takim przypadku stosuje się estymator *największej wiarygodności* ze względu na dobre własności asymptotyczne i umiarkowane wymagania obliczeniowe. Podejście to jest szczególnie wygodne dla celów predykcji, sterowania czy symulacji. Warunkiem uzyskania modelu obiektu niezależnie od sposobu pobudzania jest tu rozdzielność obiektu od źródła tego pobudzenia. Przedstawione w pracy rozważania teoretyczne ilustrowane są przykładem identyfikacji rzeczywistego obiektu elektroakustycznego, którego model jest potrzebny do prawidłowej parametryzacji układu

aktywnego tłumienia hałasu (ATH). Obiekty tego typu są szczególnie trudne do identyfikacji, gdyż cechują się bardzo złożoną dynamiką.

Końcowy wynik identyfikacji zależy od sposobu wykonania doświadczenia identyfikacyjnego. Aby móc poprawnie wnioskować na podstawie uzyskanych danych muszą być spełnione dwa podstawowe wymagania: obiekt musi być odpowiednio pobudzony oraz spełniony musi być warunek powtarzalności. Planowanie doświadczenia obejmuje także określenie wielkości mających aspekt bardziej techniczny, takich jak sprzężenie procesu i komputera. wybór okresu próbkowania czy sygnału pobudzającego. Ponieważ planowanie eksperymentu stanowi ten element procedury identyfikacji, który zależy silnie od indywidualnych cech badanego procesu oraz celu, dla którego wykonuje się identyfikację, stąd wspomaganie użytkownika w tym względzie sprowadza się głównie do dostarczenia mu zestawu porad, w jaki sposób należy prawidłowo eksperyment wykonać. Część z przedstawionych warunków stanowią założenia, których nie da się zweryfikować na drodze testowania. Jednak niektóre z nich mogą być testowane i dlatego przedstawiono przykłady testów zalecanych do wykonania po zebraniu danych. Ponieważ na ogół dane surowe nie powinny być bezpośrednio stosowane do celów identyfikacji, lecz powinny zostać odpowiednio przygotowane, stąd przedstawiono także propozycje wspomagania w tym zakresie.

Kolejnym krokiem w procedurze jest przyjęcie struktury modelu. Niech *model M* oznacza pewien system, którego wyjście przyjmuje wartości w tej samej przestrzeni, co wyjście obiektu w odpowiedzi na to samo pobudzenie. Aby móc wyznaczyć model M trzeba najpierw określić podprzestrzeń, w której M się znajduje. Powinna być ona znacznie mniejsza niż ta, w której występuje obiekt, co jest zgodne z intuicyjnie oczywistym warunkiem, że za pomocą równań matematycznych nie można opisać dokładnie procesu fizycznego oraz z tym, że dane doświadczenie może dostarczyć jedynie ograniczonej informacji. Ograniczenie przestrzeni odbywa się na drodze przyjęcia pewnej *struktury* $\mathcal{M}$, rozumianej jako zbiór wszystkich „próbnych" modeli M posiadających tę samą postać. W celu uzyskania opisu obiektu w postaci oszczędnej należy model M sparametryzować tak, aby opis ten z jednej strony uwzględniał cel, dla jakiego jest tworzony, a z drugiej własności identyfikowanego obiektu. Z uwagi na zalety powszechnie stosowanego opisu w postaci stochastycznych modeli wejściowo-wyjściowych liniowych, w dalszej części skupiono się na tej klasie modeli. Ze względu na różne sposoby modelowania zakłóceń, wśród struktur modeli o wielu wejściach (*MISO*) można wyróżnić różne typy. Do opisu obiektu elektro-akustycznego wybrany został typ *ARX*.

Tematem kolejnego rozdziału jest dyskusja szczegółów dotyczących wykonywania procedury identyfikacji. Zgodnie z metodologią eksperymentalnej metody badawczej definiuje się „prawdę" jako taką hipotezę, która nigdy nie zostanie sfalsyfikowana przez jakąkolwiek

obserwację. Ponieważ w praktyce nie jest to możliwe, stąd procedura ta w ogólności nigdy nie powinna się zakończyć. Jednak gdy model tworzony jest dla celów inżynierskich, procedura identyfikacji może zostać zatrzymana po uzyskaniu modelu, który można uznać za *odpowiedni* do tego celu. Aby powiązać to, co należy wyznaczyć (tj. model), z tym, co obserwuje się w doświadczeniu (dane), wykorzystuje się wnioskowanie statystyczne, którego celem jest wyznaczenie *wiarygodności* tej „prawdy". Identyfikacja powinna zakończyć się wybraniem jakiegoś konkretnego modelu $\hat{M}$ należącego do założonego zbioru $\mathcal{M}$. Jednocześnie model ten powinien spełniać dwa podstawowe wymagania: powinien być *odpowiedni do celu* oraz *niesprzeczny z danymi*. W klasycznym ujęciu realizowane jest to w ten sposób, że najpierw określa się pewne *kryterium aproksymacji*, które umożliwia wyznaczenie $\hat{M}$, a następnie dokonuje się jego *walidacji* i *falsyfikacji*. W rozpatrywanym przykładzie identyfikacji obiektu elektroakustycznego kryterium aproksymacji oparte jest na minimalizacji błędu predykcji, a walidacja na porównaniu charakterystyki amplitudowo-fazowej modelu parametrycznego z charakterystyką uzyskaną klasyczną metodą analizy spektralnej.

Uwzględniając powyższe uwagi przedstawiono iteracyjną procedurę identyfikacji w postaci sformalizowanej oraz wyróżniono ten fragment procedury, który może być automatycznie wykonywany przez komputer. Jednak procedura ta nie gwarantuje otrzymania poprawnego wyniku w każdej sytuacji. Zdarzają się bowiem takie przypadki, gdy identyfikacja kończy się uzyskaniem błędnego modelu, czego nie da się stwierdzić na podstawie zastosowanej procedury. Stąd omówiono możliwe przyczyny powstawania błędów oraz ewentualne inne źródła informacji. W dalszej części przedstawiono szczegóły dotyczące realizacji procedury dla modeli liniowych, takie jak kryterium wyboru, opis algorytmu poszukiwania struktury czy testowanie i analiza otrzymanego modelu. Zaproponowano procedurę automatycznego wyznaczania struktury modelu MISO minimalizującej wybrane kryterium w połączeniu z zadaniem estymacji parametrów. Zagadnienie to jest szczególnie ważne w przypadku identyfikacji procesu, którego struktura jest użytkownikowi nieznana, albo trudna do określenia *a priori*. Przedstawiono także propozycje testów wspomagających walidację i falsyfikację modelu, w szczególności analizę numeryczną uzyskanych wyników.

W komputerowo wspomaganej identyfikacji współpracują w sposób *interaktywny*: osoba prowadząca identyfikację, która dostarcza *informacji a priori* i wywołuje odpowiednie narzędzia do identyfikacji, oraz komputer wykonujący ściśle zdefiniowane zadania przetwarzania danych. Aby użytkownik mógł otrzymać potrzebne wsparcie, system komputerowego wspomagania identyfikacji powinien być zaimplementowany w odpowiedni sposób, w szczególności w zakresie interfejsu użytkownika. Niektóre z umiejętności potrzebnych w trakcie realizacji procedury identyfikacji związane są z zadaniami czysto

matematycznymi i mogą być zaprogramowanie w taki sposób, by osoba prowadząca identyfikację była inteligentnie wspomagana. Typowym przykładem jest tu zaproponowany algorytm poszukiwania struktury modelu liniowego. Natomiast w przypadku tych fragmentów procedury, które są trudne czy wręcz niemożliwe do automatycznego wykonania, wspomaganie takie powinno polegać na przedstawieniu *podpowiedzi* bazujących *na wiedzy* i *doświadczeniu projektanta* systemu, pozostawiając końcową decyzję użytkownikowi, który przejmuje za nią odpowiedzialność kierując się konkretnym zastosowaniem. Po przedstawieniu wymagań, jakim powinno sprostać oprogramowanie systemu, przeprowadzono dyskusję możliwych rozwiązań podkreślając konieczność pewnego kompromisu między elastycznością a prostotą użytkowania. Uwieńczeniem wieloletnich prac nad stworzeniem systemu wspomagającego użytkownika w identyfikacji procesów jest pakiet programowy o nazwie *MULTI-EDIP*.

Pakiet ten wykorzystano w komputerowo wspomaganej identyfikacji przykładowego obiektu elektroakustycznego w celu zaprojektowania układu kompensacji ATH do tworzenia lokalnych stref ciszy w pomieszczeniu. Ponieważ brak jest odpowiedniej wiedzy *a priori* do przyjęcia struktury modelu, stąd struktura ta wyznaczana była automatycznie przez system. Identyfikowane były zarówno modele *MISO*, jak i *SISO* po dekompozycji wykorzystującej własności wielosinusoidalnych sygnałów wektorowych ortogonalnych. Przedstawiono wyniki identyfikacji obiektu w sytuacji, gdy w trakcie eksperymentu do pomieszczenia nie docierał hałas, jak i w przypadku generowania hałasu o różnych własnościach. Dzięki zastosowaniu odpowiedniej filtracji w dziedzinie częstotliwości uzyskano znaczną poprawę wyników. Na podstawie przeprowadzonych badań można stwierdzić, że podejście polegające na przyjęciu kryterium aproksymacji opartego na minimalizacji błędu predykcji oraz użycie kryterium informacyjnego jako substytutu kryterium walidacji w iteracyjnej procedurze poszukiwania struktury modelu okazało się przydatne dla postawionego celu identyfikacji. Jednocześnie przykład ten pokazuje dobitnie, że nie można oczekiwać od systemu komputerowego wspomagania identyfikacji pełnej automatyzacji procedury identyfikacji. Potrzebna jest, szczególnie w trudniejszych przypadkach, wnikliwa analiza wyników. Jednak pomoc takiego systemu jak np. *MULTI-EDIP* jest nieoceniona.

Projektowanie systemu komputerowego wspomagania identyfikacji stwarza nowe możliwości, ale niesie także ze sobą niebezpieczeństwo powstania pewnych pułapek. Dotyczy to zarówno twórców takich systemów, jak i ich użytkowników. Stąd w ramach podsumowania omówiono perspektywy i możliwe zagrożenia w rozwijaniu oraz stosowaniu systemów komputerowego wspomagania identyfikacji.

WYDAWNICTWO POLITECHNIKI ŚLĄSKIEJ
ul. Akademicka 5, 44-100 Gliwice; tel./fax (0-32) 237-13-81
http://wydawnictwo.polsl.pl

Sprzedaż i Marketing
tel. (0-32) 237-18-48
wydawnictwo_mark@polsl.pl

Nakł. 100+50	Ark. wyd. 13	Ark. druk. 11,375	Papier offset. 70x100,80g
Oddano do druku 12.06.2006.r.	Podpisano do druku 12.06.2006 r.		Druk ukończ. w czerwcu 2006 r.

Wydrukowano w Zakładzie Graficznym Politechniki Śląskiej w Gliwicach, ul. Kujawska 1
zam. 211/06

Książki Wydawnictwa można nabyć w księgarniach

GLIWICE

- Punkt Sprzedaży Wydawnictwa na Wydziale Górnictwa i Geologii – ul. Akademicka 2 (237-17-87)
- „FORMAT" – Akademicka 5 na Wydziale Budownictwa
- „LAMBDA" – ul. Akademicka 2 (237-21-40)
- „MERCURIUS" – ul. Prymasa S.Wyszyńskiego 14 b (032) 230-47-22
- „ŻAK" – ul. Kaszubska (budynek Biblioteki)

BIAŁYSTOK

- Dom Książki (Księgarnia 84) – ul. Wiejska 45 c
- EKOPRESS Księgarnia Wysyłkowa - ul. Brukowa 28 (085) 746-04-95

GDAŃSK

- EKO-BIS – ul. Dyrekcyjna 6 (058) 305-28-53

KATOWICE

- Punkt Sprzedaży na Wydziale Transportu – ul. Krasińskiego 8
- Hurtownia „DIK" – ul. Dulęby 7 (032) 204-82-30
- Hurtownia „JERZY" – ul. Słoneczna 24 (032) 258-99-58

KRAKÓW

- Techniczna – ul. Podwale 4 (012) 422-48-09
- Punkt Sprzedaży WND – AGH, Al. Mickiewicza 30 (012) 634-46-40

ŁÓDŹ

- „POLITECHNIKA 100" – ul. Żeromskiego 116 PŁ.
- Hurtownia „BIBLIOFIL" – ul. Jędrowizna 9a (042) 679-26-77

OPOLE

- BK - „POLITECHNIKA" – Wydz. Budownictwa, ul. Katowicka 48 (077) 456-50-58 wew.333

POZNAŃ

- Księgarnia „POLITECHNIK" – ul. Piotrowo 3 (061) 665-23-24
- Księgarnia Techniczna – ul. Półwiejska 28 (061) 659-00-38

RYBNIK

- „ORBITA" – ul. Rynek 12
- „NEMEZIS" – ul. Hallera 26

TYCHY

- „I JA TOURS" - ul. Piłsudskiego 10 (217-00-91 w.130)

WARSZAWA

- Studencka – Pl. Politechniki 1 (022) 628-77-58
- Techniczna - ul. Kaliskiego 15 (022) 666-98-02
- Techniczna – ul. Świętokrzyska 14
- MDM – ul. Piękna 31

WROCŁAW

- „TECH" – ul. Wybrzeże Wyspiańskiego 27

ZABRZE

- Punkt Sprzedaży na Wydziale Organizacji i Zarządzania– ul. Roosevelta 26

BIBLIOTEKA GŁÓWNA
Politechniki Śląskiej
P.3341/06/142

Druk: Drukarnia Gliwice, ul. Zwycięstwa 27, tel. 230 49 60