

Witold Malina, Mariusz Szwoch



PODSTAWY PROJEKTOWANIA INTERFEJSÓW UŻYTKOWNIKA

Helion 

Wszelkie prawa zastrzeżone. Nieautoryzowane rozpowszechnianie całości lub fragmentu niniejszej publikacji w jakiegokolwiek postaci jest zabronione. Wykonywanie kopii metodą kserograficzną, fotograficzną, a także kopiowanie książki na nośniku filmowym, magnetycznym lub innym powoduje naruszenie praw autorskich niniejszej publikacji.

Wszystkie znaki występujące w tekście są zastrzeżonymi znakami firmowymi bądź towarowymi ich właścicieli.

Autor oraz Wydawnictwo HELION dołożyli wszelkich starań, by zawarte w tej książce informacje były kompletne i rzetelne. Nie biorą jednak żadnej odpowiedzialności ani za ich wykorzystanie, ani za związane z tym ewentualne naruszenie praw patentowych lub autorskich. Autor oraz Wydawnictwo HELION nie ponoszą również żadnej odpowiedzialności za ewentualne szkody wynikłe z wykorzystania informacji zawartych w książce.

Redaktor prowadzący: Małgorzata Kulik

Projekt okładki Studio Gravite / Olsztyn
Obarek, Pokoński, Pazdrijowski, Zaprucki

Grafika na okładce została wykorzystana za zgodą Shutterstock.com

Wydawnictwo HELION
ul. Kościuszki 1c, 44-100 GLIWICE
tel. 32 231 22 19, 32 230 98 63
e-mail: helion@helion.pl
WWW: <http://helion.pl> (księgarnia internetowa, katalog książek)

Drogi Czytelniku!

Jeżeli chcesz ocenić tę książkę, zajrzyj pod adres
<http://helion.pl/user/opinie/poprin>
Możesz tam wpisać swoje uwagi, spostrzeżenia, recenzję.

ISBN: 978-83-283-3790-9

Copyright © Helion 2018

Printed in Poland.

- [Kup książkę](#)
- [Poleć książkę](#)
- [Oceń książkę](#)

- [Księgarnia internetowa](#)
- [Lubię to! » Nasza społeczność](#)

Spis treści

Wstęp	5
Rozdział 1. Wprowadzenie do interfejsów użytkownika	7
1.1. Interfejs użytkownika	7
1.2. Komputer i jego otoczenie	10
1.3. Historia rozwoju interfejsów człowieka z komputerem	11
1.4. Interfejs dla osób niepełnosprawnych	16
Rozdział 2. Urządzenia interakcji	19
2.1. Urządzenia wejściowe	20
2.2. Urządzenia wyjściowe	27
2.3. Uwagi końcowe	31
Rozdział 3. Style interakcji człowieka z komputerem	33
3.1. Wprowadzenie	33
3.2. Interfejs tekstowy i języki poleceń	34
3.3. Graficzny interfejs użytkownika	36
3.3.1. Charakterystyka GUI	37
3.3.2. Kategorie kontrolek	37
3.4. Menu	39
3.4.1. Menu pojedyncze i liniowe	41
3.4.2. Menu drzewiaste i grafy menu	43
3.4.3. Menu stałe i kontekstowe	50
3.4.4. Uwagi końcowe	55
3.5. Okna	58
3.5.1. Struktura i zawartość okna	58
3.5.2. Rodzaje okien	61
3.5.3. Zarządzanie wieloma oknami	63
3.6. Ikony i kafelki	69
3.6.1. Ikony	69
3.6.2. Kafelki	72
3.7. Manipulacja bezpośrednia	73
3.7.1. Kursory	74
3.7.2. Zastosowania manipulacji	75
3.7.3. Wady i zalety manipulacji bezpośredniej	75

3.8. Formularze	78
3.8.1. Numer telefonu	81
3.8.2. Daty	81
3.8.3. Czas	81
3.8.4. Waluta	82
3.8.5. Uwagi końcowe	82
Rozdział 4. Systemy multimedialne i multimodalne	83
4.1. Multimedia — rozwój i zastosowanie	83
4.2. Sygnały akustyczne i język naturalny	86
4.2.1. Interfejsy dźwiękowe	86
4.2.2. Systemy informacji głosowej	88
4.2.3. Generacja (synteza) mowy	88
4.2.4. Rozpoznawanie słów	89
4.2.5. Mowa ciągła	89
4.3. Gesty	91
4.4. Uwagi końcowe	91
Rozdział 5. Projektowanie interfejsu człowiek – komputer	93
5.1. Wprowadzenie	93
5.2. Analiza zadań i użytkowników	94
5.3. Projektowanie struktury interfejsu	95
5.4. Wybór urządzeń interfejsu i stylu interakcji	98
5.5. Prezentacja informacji wyjściowej	100
5.5.1. Diagram Gutenberga	100
5.5.2. Projektowanie ekranów aplikacji	101
5.5.3. Wybór kolorów	103
5.6. Formy pomocy dla użytkownika	104
5.7. Prototypowanie i ocena interfejsu	105
5.7.1. Narzędzia prototypowania graficznego interfejsu użytkownika	106
5.7.2. Ocena interfejsu	110
5.7.3. Uwagi końcowe	110
Rozdział 6. Rzeczywistość wirtualna i rozszerzona	113
6.1. Rzeczywistość wirtualna	113
6.1.1. Interfejs i interakcja	115
6.1.2. Zastosowania VR	118
6.2. Rzeczywistość rozszerzona	122
6.3. Uwagi końcowe	126
Zakończenie	129
Literatura	131
Źródła zdjęć	135
Skorowidz	137

Rozdział 4.

Systemy multimedialne i multimodalne

Dzięki nieustannemu udoskonalaniu sprzętu komputerowego zwiększają się jego możliwości współpracy z różnymi urządzeniami cyfrowymi, co pozwala na wzbogacanie metod interakcji człowieka z komputerem i w efekcie rozwoju interfejsu aplikacji. Ten kierunek rozwoju urzeczywistnia się w postaci systemów multimodalnych i w ten sposób wymiana informacji z komputerem może angażować wiele kanałów percepcji użytkownika.

4.1. Multimedia — rozwój i zastosowanie

Systemy multimedialne stanowią kolejny wielki etap rozwoju systemów komputerowych PC, wyrażający się szeroką integracją urządzeń peryferyjnych z komputerem. Twórcy dążą do tego, by nowe aplikacje mogły używać dźwięku, obrazu i wideo jako części standardowego interfejsu użytkownika i by zostały zaakceptowane tak samo, jak miało to miejsce w przypadku interfejsów graficznych, które zastąpiły używany wcześniej tryb tekstowy. Główne kierunki zastosowania multimediów są opisane niżej [1, 2, 25, 34, 38].

Pierwsze systemy multimedialne bazowały na idei hipertekstu i były systemami generalnie tekstowymi, wzbogaconymi o obrazy cyfrowe i ewentualnie dźwięk. Ich atrakcyjność polegała głównie na powiązaniu zgromadzonych informacji i ciekawej prezentacji multimedialnej. Szybki rozwój urządzeń multimedialnych, takich jak karty graficzne, karty dźwiękowe, wideo, CD-ROM itp., przyczynił się do istotnego poszerzenia funkcji w zakresie przetwarzania, przechowywania i wyszukiwania informacji multimedialnej. Koncepcja programowania obiektowego oraz systemy operacyjne wysokiego poziomu umożliwiły projektowanie prostych i łatwych w użyciu systemów

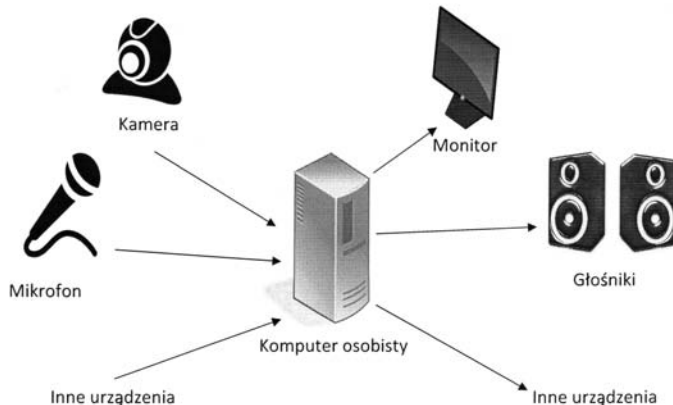
multimedialnych. Stało się przy tym możliwe przetwarzanie danych w czasie rzeczywistym oraz samodzielne sterowanie programem przez użytkownika, tj. praca interaktywna.

Wytwarzanie aplikacji multimedialnych, zwłaszcza opartych na niestandardowych urządzeniach wejściowych i wyjściowych, jest zadaniem złożonym, w którym należy wziąć pod uwagę czynniki takie jak różnorodność sprzętu i oprogramowania oraz użytkowników, a także rozmaite warunki użytkowania, co jest istotne przy korzystaniu z optycznego lub akustycznego wejścia do systemu. Wytworzenie niezawodnego i uniwersalnego oprogramowania, wykorzystującego cały potencjał sprzętu, wymaga stosowania nowoczesnych rozwiązań i metodologii wytwarzania oprogramowania, które kładą duży nacisk na etapy testowania, co z kolei znacząco podnosi koszt całego procesu. Powoduje to, że rozwój nowatorskich technologii i urządzeń, które mogłyby nadać programom multimedialnym nowy wymiar (takich jak interfejsy trójwymiarowe, akcesoria dla rzeczywistości wirtualnej czy rozpoznawanie mowy), przebiega stosunkowo wolno.

Obecnie właściwie każdy nowy zestaw komputerowy zasługuje na miano multimedialnego, ponieważ już standardem są karty graficzne wyświetlające kolorowy obraz wysokiej rozdzielczości, karty dźwiękowe i dyski optyczne. Ponadto taki komplet może zostać wyposażony w kamerę, mikrofon oraz wzmacniacz z kolumnami głośnikowymi (rysunek 4.1) [36]. Krótko mówiąc, typowy zestaw multimedialny to taki, którego aplikacje mogą posługiwać się tekstem, dźwiękiem, obrazem i wideo.

Rysunek 4.1.

Przykład konfiguracji multimedialnej komputera osobistego

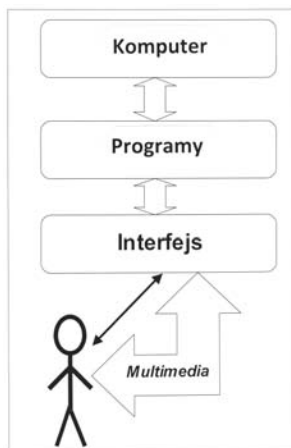


Miejsce interfejsu w systemie multimedialnym pokazano na rysunku 4.2.

System multimedialny musi być traktowany przez projektantów z jednej strony jako narzędzie do przetwarzania informacji w konkretnym celu (np. do wspomagania decyzji z zakresu marketingu, finansów czy w grach), z drugiej zaś jako medium, które zachęci użytkownika do korzystania z aplikacji (tutaj ważna jest estetyka, łatwość komunikacji itp.). Aplikacja multimedialna może być prezentacją pasywną, porównywalną z programem telewizyjnym, lub też interaktywnym systemem pozwalającym użytkownikowi na sterowanie nieliniowym przebiegiem wykonania zadania.

Rysunek 4.2.

*Interfejs i multimedia
w dialogu człowieka
z komputerem*



Obok branży rozrywkowej to edukacja stanowi największy rynek zbytu systemów multimedialnych, najlepiej bowiem wykorzystuje ich istotę, czyli możliwość prezentacji zagadnień. Dobrze znane są tu przykłady licznych programów do nauki języków obcych, encyklopedie multimedialne czy też multimedialne wersje podręczników szkolnych i filmów do różnych przedmiotów. Dzięki ich zastosowaniu w szkołach można ciekawiej prowadzić lekcje, uczniowie zaś mogą łatwiej rozwiązywać zadania i rozwijać własne zainteresowania pogłębiające ich wiedzę. Z drugiej strony rynek zalewany jest przez dziesiątki pozycji wykorzystujących co prawda technologie multimedialne, lecz z punktu widzenia edukacyjnego bezużytecznych lub wręcz zawierających merytoryczne błędy. Zmusza to nauczycieli do szczególnie dokładnej selekcji oprogramowania, zwłaszcza że zakup pojedynczej kopii programu komputerowego może oznaczać spory wydatek. Natomiast podzielany powszechnie pogląd, że komputer multimedialny podłączony do internetu stanowi uniwersalną receptę na podniesienie poziomu kształcenia, jest z pewnością błędny.

Z uwagi na wysokie koszty wytwarzania i użytkowania aplikacji multimedialnych ich rozwój ma sens, jeżeli poniesione wydatki zostaną zrekompensowane poprzez dużą liczbę sprzedanych licencji, jak ma to miejsce w sektorze edukacji i rozrywki (ang. *edutainment* = *education*+*entertainment*), lub poprzez inne wymierne korzyści odbiorcy biznesowego, np. redukcję czasu szkolenia czy zmniejszenie liczby personelu przy jednoczesnym zagwarantowaniu jakości i efektywności szkolenia. Świat biznesu i administracji docenia efektywność działania systemów multimedialnych. Na przykład baza danych skupiająca pracowników administracji może być znacznie poszerzona i oprócz pól tekstowych może zawierać zdjęcia i nagrania głosu poszczególnych osób. Pojawiły się także komputerowe systemy wspomagania decyzji, a coraz większą popularność zdobywają multimedialne systemy identyfikacji i kontroli wykorzystujące techniki rozpoznawania obrazu i dźwięku, niekiedy dysponujące możliwością analizy nagrań głosu i wideo. Systemy multimedialne znajdują obecnie zastosowanie w wielu dziedzinach życia. Można tu wymienić m.in.: systemy do śledzenia obiektów i monitoringu w miejscach publicznych (na ulicy, lotniskach, dworcach, w sklepach itp.) czy do opieki nad osobami starszymi i niepełnosprawnymi. Zastosowanie kamery termowizyjnej na podczerwień stwarza dodatkowe możliwości wykorzystania takich systemów przez wojsko lub służbę zdrowia.

4.2. Sygnały akustyczne i język naturalny

Użycie języka naturalnego we współdziałaniu z komputerem jest postrzegane jako najbardziej pożądaný styl interakcji, pozwalający użytkownikowi komunikować się z systemem w sposób swobodny. Mowa jest szybką metodą komunikacji i doskonale sprawdza się w sytuacjach, w których użytkownik nie może wydawać poleceń za pomocą innych form interakcji. Język naturalny jest odpowiedni dla każdego użytkownika, chociaż przeprowadzany za jego pomocą dialog wcale nie musi być szybszy.

Nadzieja, że komputer będzie reagował na polecenia i odpowiadał na pytania wydawane głosem, nieustannie inspiruje naukowców i inżynierów. W ostatnim czasie zanotowano ogromny postęp w zakresie rozpoznawania i generowania mowy. Zastosowanie praktyczne głosu w interakcji ma jednak sens tylko wtedy, kiedy użytkownicy mogą działać szybko i z małym poziomem błędu [2].

Projektowanie z użyciem interakcji audio obejmuje przynajmniej pięć odmian:

- ◆ systemy z sygnałami akustycznymi,
- ◆ systemy informacji głosowej,
- ◆ generacja (synteza) mowy,
- ◆ rozpoznawanie słów,
- ◆ rozpoznawanie mowy ciągłej.

W rozpoznawaniu słów i mowy wyróżniamy dwa rodzaje systemów:

- ◆ systemy mowy zależnej (od mówcy),
- ◆ systemy mowy niezależnej.

Wymienione grupy mogą być łączone na różne sposoby i prowadzić do różnych systemów.

4.2.1. Interfejsy dźwiękowe

Dźwięk ma duże znaczenie w przekazywaniu informacji. Każdy dźwięk charakteryzuje się swoją częstotliwością, siłą (głośnością) i tembrem (barwą). Człowiek słyszy dźwięki w zakresie częstotliwości od 20 Hz do 15 kHz. Do zaprojektowania interfejsu można użyć różnych dźwięków i sygnałów mowy. Wyszukiwanie i badanie wymyślnych form prezentacji informacji często odnosi się do audiowizualizacji lub interfejsu dźwiękowego. Na przykład wczesne dalekopisy generowały dźwięk dzwonka, aby ostrzec użytkownika, że skończył się papier lub poinformować, że przyszła nowa wiadomość. Później systemy komputerowe dodały różne dźwięki do wskazywania zagrożenia lub sygnalizowania zakończenia pewnej akcji. Niektóre urządzenia, takie jak kamera cyfrowa, gene-

rują elektroniczne dźwięki potwierdzające daną akcję (jako sprzężenie zwrotne). Z drugiej strony po kilku godzinach pracy takie dźwięki mogą raczej zakłócać, niż pomagać, szczególnie w pomieszczeniu z kilkoma urządzeniami i użytkownikami.

Dźwięki możemy podzielić na takie, które występują w sposób naturalny w środowisku, lub abstrakcyjne — generowane przez sztuczne źródło. W pierwszym przypadku mogą to być dźwięki takie jak skrzypienie drzwi, przelewanie płynu lub dzwonienie, które pomagają wzmocnić wizualne metafory zastosowane w graficznym interfejsie. W dźwiękach naturalnych rozpoznaje się nie tylko samo źródło, ale i jego charakter.

W drugim przypadku do reprezentacji obiektów i akcji są używane ikony dźwiękowe (ang. *auditory icons*). Na przykład „katalog” oznacza zwykle zbiór kartek papieru, dlatego podczas jego likwidowania (przenoszenia do kosza) emitowany jest dźwięk niszczonego papieru. W innej sytuacji efekt dźwiękowy można utworzyć, ale jego znaczenia trzeba się nauczyć. Na przykład kopiowanie nie ma przypisanego dźwięku, ale może być sygnalizowane przez odgłos przelewania płynu do pojemnika, przy czym wysokość towarzyszącego dźwięku rośnie wraz z postępem kopiowania. Dźwięki mogą przekazywać wiele informacji bardzo ekonomicznie. Gdy do skrzynki pocztowej nadejdzie plik — usłyszymy szelest, gdy plik jest duży — wystąpi niski, ciężki dźwięk, jeżeli będzie to plik tekstowy — szmer, a gdy skompilowany program — dźwięk metalu.

Różnorodność użytych sygnałów i dźwięków może być atrybutem menu akustycznego, właściwego w sytuacji, gdy użytkownik np. prowadzi samochód i ma zajęte ręce i oczy. Menu akustyczne jest także użyteczne w interfejsie telefonicznym i w rozmaitych informatorach powszechnego dostępu. W menu akustycznym odpowiedzi i listy opcji są wypowiadane do użytkownika, który odpowiada, używając stosownych przycisków klawiatury, ekranu dotykowego albo własnego głosu.

Ikony dźwiękowe możemy podzielić na kilka kategorii ze względu na funkcje, jakie pełnią:

- ♦ zawiadomienie użytkownika o konkretnym zdarzeniu, np. pojawieniu się nowej poczty,
- ♦ potwierdzenie efektów działań użytkownika, np. zapisania lub skasowania pliku,
- ♦ sygnalizacja zmian stanu, wynikająca z działania użytkownika,
- ♦ wskazywanie użytkownikowi wybranych obiektów,
- ♦ proponowanie odpowiedzi (w rodzaju sugerowanych odpowiedzi *tak* lub *nie*),
- ♦ sygnalizacja działania w toku (monitorowanie działania).

W ten sposób aplikacje mogą skutecznie i w przewidziany sposób wpływać na działania użytkownika. Emisja sygnałów dźwiękowych trwa krótko, zwykle od 0,1 do 2 sekund.

Rozwiązania akustyczne mogą wykorzystywać indywidualne dźwięki, a bardziej złożone informacje prezentować przez kombinację dźwięku i muzyki. Od projektantów dźwięku oczekuje się także uczestnictwa w nowych projektach, szczególnie mobilnych.

4.2.2. Systemy informacji głosowej

Znaczenie języków naturalnych w nowoczesnych systemach telefonicznych i informacyjnych stale rośnie. Potrafią one wspierać użytkownika w realizacji niektórych zadań. Na przykład podczas próby połączenia telefonicznego z pożądanym departamentem czy osobą w danej organizacji system, używając nagranej wiadomości, udostępnia użytkownikowi kilka ponumerowanych możliwości wyboru. Użytkownik wskazuje opcję, wypowiadając lub naciskając odpowiedni numer na klawiaturze telefonu. Następnie system może przedstawić kolejny zbiór opcji. Proces ten trwa do momentu zebrania informacji koniecznych do zrealizowania połączenia. Powyższy przykład pokazuje, że nawet system posługujący się bardzo ograniczonym słownictwem może być przydatny.

Systemy informacji głosowej są popularne, ponieważ umożliwiają obsługiwanie klientów przez całą dobę i nie są drogie. Mają one także szerokie zastosowanie w nawigacji samochodowej, serwisach internetowych, pomieszczeniach dyspozytorskich i kontrolnych (np. komunikat typu: „Uwaga, podnosi się temperatura!”), poczcie głosowej, grach dla dzieci itp.

4.2.3. Generacja (synteza) mowy

Generacja mowy jest niekwestionowanym sukcesem naukowym i technologicznym, co więcej — ma wiele zastosowań praktycznych. Syntezator mowy może być użyty do rozszerzenia graficznego interfejsu użytkownika, a w sieciach Web do tworzenia przeglądarek głosowych. Umożliwia on czytanie na głos dokumentów, gazet czy danych statystycznych. Możliwość zamiany tekstu na głos jest bardzo atrakcyjna zwłaszcza dla niewidomych. Takie rozwiązanie jest zaimplementowane w interfejsie Microsoft Windows Narrator oraz w sieci Web w przeglądarkach głosowych (ang. *voice browsers*) do czytania fragmentów tekstu. Dobrze gdy generowane wiadomości są proste i krótkie.

Jednym z ciekawych osiągnięć z tego obszaru jest polski syntezator mowy IVONA¹. Spółka Unikon Integral opracowała w 2009 r. program MagicScribe, który rozpoznaje mowę i zamienia ją na tekst. Z kolei program opracowany specjalnie dla lekarzy pozwala im poświęcać dużo mniej czasu na tworzenie dokumentacji (program jest bowiem trzy razy szybszy niż przeciętna osoba pisząca na komputerze). Opracowano też podobne programy, które służą prawnikom, administracji, firmom, a nawet studentom, którzy chcą automatycznie tworzyć notatki. Dalsze badania koncentrują się na użyciu tej metody w telekomunikacji w celu odczytu SMS-ów i uwolnienia użytkownika od męczącego odpisywania na małej klawiaturze. Duże znaczenie ma opracowanie syntezatorów dla urządzeń mobilnych, które to stają się na rynku ważniejsze od komputerów osobistych.

¹ <http://www.ivona.com/pl/>.

Firma IVO Software opracowała kolejną, ulepszoną wersję syntezatora mowy — IVONA 2. Nowy syntezytor jest bardziej inteligentny. Potrafi generować pojedyncze słowa i całe zdania oraz czytać książki z jakością zbliżoną do ludzkiej mowy, ponadto jest dziesięciokrotnie szybszy od poprzedniego. W języku polskim oferuje cztery rodzaje głosów, a w języku angielskim sześć.

4.2.4. Rozpoznawanie słów

Obecnie automatyczne rozpoznawanie izolowanych słów osiągnęło dość wysoki poziom, co otwiera przed nim szerokie pole zastosowania. Współczesne urządzenia rozpoznają słowa wypowiedziane przez konkretną osobę z niezawodnością 90 – 98% dla zbioru 100 – 1000 słów lub nawet większych. Rozpoznawanie słów funkcjonuje dobrze w wyspecjalizowanych zastosowaniach — w wojsku, lotnictwie, w medycynie (np. na sali operacyjnej), w laboratoriach i biurach — ale nie jako ogólny środek interakcji.

Podczas codziennego korzystania z komputera, kiedy używa się monitora z ekranem, głos ma marginalne znaczenie. Klawiatura i urządzenie wskazujące z bezpośrednią manipulacją są często szybsze w działaniu niż rozpoznawanie mowy. Chociaż kombinacja głosu i bezpośredniej manipulacji może być użyteczna, to polecenia błędnie podane na wejście głosowe mogą być trudne do naprawy, a i efektywna reakcja systemu może być zbyt wolna.

Problemy pojawiają się, kiedy zmienia się poziom natężenia głosu, użytkownik jest chory lub zestresowany albo słowa są do siebie podobne. Projekty badawcze skupiają się na poprawie jakości rozpoznawania głosu w trudnych warunkach, zwiększeniu podstawowego słownika do ponad 10 tys. słów, ograniczeniu czasu rozpoznawania przy rozszerzeniu zasobu słów i eliminacji konieczności nauczania się przez urządzenie specyficznego głosu poszczególnych użytkowników. System mowy zależnej od mówcy daje większą dokładność niż system z mówcą niezależnym, ale ten drugi eliminuje etap treningu i w ten sposób rozszerza zakres komercyjnych zastosowań.

4.2.5. Mowa ciągła

Badania naukowe poświęcane rozpoznawaniu mowy ciągłej prowadzone są z nadzieją na komercyjny sukces. Jednak użytkownicy są często rozczarowani poziomem błędów i trudnością ich poprawienia. Projektantom programów komputerowych kłopot sprawiają zwykłe rozmyte granice między wypowiedzianymi słowami. Do innych problemów należą: różnorodność akcentów, odmienne poziomy wypowiedziania, wpływ emocji na intonację oraz skala zakłóceń i szumów. Jednak większość błędów wynika z semantycznej interpretacji i braku kontekstowego rozumienia, które człowiek łatwo wykorzystuje w celu wyeliminowania niejednoznaczności. Język naturalny silnie zależy od kontekstu, przez co często jest dwuznaczny.

Systemy rozpoznawania mowy ciągłej umożliwiają użytkownikowi na przykład dyktowanie tekstów i składanie raportów słownych do automatycznej translacji. Przeglądanie i poprawianie ich na komputerze odbywa się zwykle na ekranie przy użyciu klawiatury. Zastosowanie metod sztucznej inteligencji pozwala na dalsze ułatwienia w zakresie poprawiania tekstu oraz automatycznego opracowania streszczenia.

Oprócz rozpoznawania treści w niektórych sytuacjach zachodzi potrzeba automatycznego rozpoznawania głosu (czyli mówcy). Jest to możliwe, ponieważ w sygnale mowy poza informacjami określającymi treść wypowiedzi są zawarte również informacje osobnicze charakteryzujące indywidualny głos mówcy. Tego rodzaju cechy osobnicze głosu zawarte w sygnale mowy umożliwiają rozpoznanie mówcy na podstawie jego wypowiedzi.

Rozpoznawanie mowy może być użyte do identyfikacji osób. Spotykamy się z tym na co dzień, gdy musimy rozpoznać głos drugiej osoby podczas rozmowy telefonicznej lub głos aktora w audycji radiowej. Zastosowanie takiej identyfikacji w systemach komputerowych może umożliwiać uprawnionym osobom dostęp do komputera w celu korzystania z jego oprogramowania i baz danych. Innym ważnym obszarem wykorzystania rozpoznawania głosu jest bankowość (np. przy weryfikacji klientów podczas transakcji bankowych, kontroli telefonicznego dostępu do kont bankowych czy podczas korzystania z kart kredytowych), ale także wojsko (np. do zwiększenia bezpieczeństwa). W przypadku takiego systemu wskazane jest, aby dokonał on identyfikacji mówcy już po kilku wypowiedzianych słowach. Pomimo dużego zainteresowania oraz ciągłego postępu technicznego, na razie możliwości praktycznego zastosowania rozpoznawania mowy ciągłej są dość ograniczone.

Automatyczne rozpoznawanie głosów ma głównie dwa cele: identyfikację głosów oraz ich weryfikację. Identyfikacja głosów to procedura przypisania nieznanej wypowiedzi do mówcy z ustalonego zbioru. Inaczej mówiąc, w populacji M mówców należy wskazać tego, którego cechy osobnicze głosu są najbardziej zbliżone do cech głosu zawartych w analizowanej wypowiedzi. Weryfikacja głosów jest natomiast procedurą potwierdzenia lub odrzucenia przez przedstawiającego się mówcę zgłoszonej tożsamości na podstawie próbki mowy.

Systemy autoryzacji głosowej są jeszcze dość zawodne, ponieważ cechy ludzkiego głosu mogą się zmieniać w znaczący sposób — nawet z godziny na godzinę. Jednak duże znaczenie praktyczne porozumiewania się z komputerem za pomocą głosu powoduje, że prowadzone są intensywne badania nad opracowaniem skutecznych procedur rozpoznawania mowy i mówców.

Zastosowanie rozpoznawania mowy w menu drzewiastym może być sukcesem, jeżeli użytkownik zna interesujące go nazwy. Gorzej, kiedy struktura menu się rozszerza i wydłuża lub kiedy segmenty odbieranej informacji słownej są rozwlekłe — to frustruje użytkownika. Opóźnienie pojawiania się głosu na wyjściu, jego ulotna natura i ewentualne trudności w zrozumieniu pozostają dużym wyzwaniem. Pomimo występujących ograniczeń systemy tego rodzaju oferują duże korzyści dla szerokiego grona osób niepełnosprawnych, które swoje polecenia mogą wydawać właśnie głosem.

W dobie intensywnego wykorzystania i rozwoju interfejsów graficznych obserwuje się ewolucję wszystkich interfejsów użytkownika w kierunku interfejsu multimedialnego, w którym do zaprezentowania informacji zostają połączone dane różnych typów, takie jak tekst, dźwięk, obraz statyczny i dynamiczny, czyli wideo.

4.3. Gesty

Ciągły rozwój technologii sprzętu komputerowego pozwala na wprowadzanie nowych form komunikacji między człowiekiem i komputerem [4, 28]. Ważną nową formą przekazywania informacji jest analiza zachowania człowieka, tj. jego ciała, a w szczególności gestów wyrażanych za pomocą ruchu i pozycji ręki, dłoni i kciuków. Osobliwość tej metody sprawia, że sprawdza się ona w pewnych sytuacjach, w których inne sposoby interakcji mogą nie być odpowiednie lub efektywne, np. wśród ludzi upośledzonych.

Ogólnie gesty ręki możemy klasyfikować jako statyczne i zmienne w czasie. Gesty statyczne reprezentują znaki izolowane, podczas gdy druga grupa wyraża sekwencje podlegające zmianom w czasie. Rozpoznawanie znaków ciągłych (podobnie jak mowy ciągłej) jest o wiele trudniejsze, ponieważ między gestami nie ma wyraźnych przerw. Przy rozpoznawaniu języka znaków i mowy ciągłej występują zatem podobne problemy, ponieważ oba procesy są zmienne w czasie.

Według przeznaczenia gesty możemy podzielić na:

- ♦ konwersacyjne,
- ♦ sterujące,
- ♦ służące do manipulacji,
- ♦ komunikacyjne.

Gesty *konwersacyjne* są wyrażane za pomocą języka znaków, który charakteryzuje się specjalnym zbiorem słów i gramatyką. Słowa w tym języku ujawniane są w postaci gestów (znaków), które mogą być (w pewnym stopniu) zmienne podczas dialogu.

Przykładem gestów *sterujących* są gesty nawigacyjne, gdzie wychwytywana jest orientacja ręki, która wskazuje kierunek w przestrzeni (w środowisku rzeczywistym lub wirtualnym) do nawigacji. Gesty *manipulacyjne* stanowią sposób interakcji człowieka z obiektami wirtualnymi (rysunek 3.37), natomiast gesty *komunikacyjne* obejmują specjalne przypadki, które analizowane są na studiach psychologicznych.

Z powyższego opisu widać, że ewolucja interfejsu użytkownika od formy tekstowej, klawiatury i myszki zrobiła ogromny postęp, a rozpoznawanie gestów, chociaż wciąż jest w początkowej fazie rozwoju, nabrało już dużego znaczenia praktycznego. Wykorzystanie do sterowania komputerem gestów, odręcznego pisma i głosu stanowi ważny etap postępu interfejsu multimodalnego.

4.4. Uwagi końcowe

Rozszerzeniem interfejsu jednomodalnego, np. interfejsu WIMP, opartego na metaforach, jest wprowadzenie interfejsu multimodalnego. Interfejs multimodalny jest połączeniem (kombinacją) różnych kanałów informacji (tekstowego, akustycznego, wizualnego),

co umożliwia różne sposoby interakcji, np. za pomocą śledzenia gałek ocznych, czytania z ruchu warg, rozpoznawania twarzy, pisma itp.

Interfejs użytkownika stał się jednym z najważniejszych elementów oprogramowania. Jego szczególne znaczenie wynika z faktu, że użytkownik nie widzi innej części systemu poza jego interfejsem. Mimo że ukuto określenie „przyjazny dla użytkownika” (ang. *user friendly*) i przyjęto, że sterowanie komputerem powinno opierać się na grafice oraz być zorientowane na człowieka, i mimo że interfejs graficzny udoskonalano przez wiele lat — nadal praca z komputerem nie dla wszystkich użytkowników jest łatwa. Jef Raskin zwraca uwagę, że obecne systemy stają się coraz większe i coraz bardziej skomplikowane oraz że pomimo ciągłych ulepszeń ich interfejsów wciąż używa się pojęć bardzo specyficznych (np. „system plików” czy „otoczenie sieciowe”), których zrozumienie stanowi podstawę sprawnej pracy z komputerem. Systemy mają również wiele ustawień parametrów i właściwości, których poznanie i wykorzystanie przez przeciętnego użytkownika nie jest proste.

Do powstania przyjaznego interfejsu może przyczynić się budowa systemów o większej inteligencji. Innym kierunkiem rozwoju interfejsu jest np. stworzenie użytkownikowi możliwości posługiwania się językiem naturalnym. Zapytania w języku naturalnym są dla użytkownika dużo łatwiejsze niż zapytania w wymyślonych do tego celu językach formalnych. Na wyższy poziom mogą wznieść interfejs użytkownika nowe urządzenia wejścia-wyjścia, jeszcze lepiej wykorzystujące głos, odręczne pismo czy nawet myśli.

Koncepcje łączące najnowsze rozwiązania sprzętowe z osiągnięciami sztucznej inteligencji stanowią obecnie podstawę dalszego rozwoju badań nad interfejsem. Intelligentny multimodalny interfejs powinien zwiększyć możliwości interakcji użytkownika z komputerem. Zastosowanie odpowiednich mediów pozwoli zaś na poszerzenie strumienia informacji przepływającego między człowiekiem a maszyną. Aby osiągnięcie takich korzyści stało się w pełni możliwe, konieczne jest zrozumienie charakteru informacji i jej powiązania z poszczególnymi mediami.

Korzyścią interakcji multimodalnej jest to, że człowiek może komunikować się w inny sposób niż tradycyjny oraz że komputer może lepiej zrozumieć polecenie, jakie powinien wykonać. Jest to szczególnie ważne w sytuacji, kiedy nie nastąpiło dokładne rozpoznanie instrukcji wydanej za pomocą gestu lub głosu. Od wielu lat trwają prace nad rozwojem interfejsów multimedialnych i multimodalnych pozwalających na wykorzystywanie różnych kanałów interakcji z komputerem.

Interfejsy tego typu pozwalają m.in. na zdalne sterowanie komputerem w sytuacji, gdy użytkownik nie może zarządzać nim bezpośrednio. Jednym z przykładów możliwości podejścia multimodalnego jest sterowanie programami komputerowymi za pomocą całego ciała, gestów lub głosem, tak jak w interfejsie typu NUI (ang. *natural user interface*) [14] stosowanym obecnie w konsoli Microsoft Xbox oraz w nowoczesnych telewizorach.

Skorowidz

A

aplikacje RAD, 106
AR, augmented reality, 113
 external, 125
 internal, 125
awatar, 117

B

BCI, brain-computer interface, 17
belka tytułowa, 58

C

CAVE, 30, 118, 119
charakterystyka stylów interakcji, 99
CLI, command-line interface, 12, 34
czarodziej, 14
czas, 81
czcionka, 79

D

dane tabelaryczne, 39
daty, 39, 81
dekompozycja funkcjonalna zadania, 96
diagram Gutenberga, 100
dostęp do sieci komputerowej, 11
doznanie akustyczne, 116
drukarka 3D, 27
drzewo, 43
 katalogów, 47
dżojstik, 24, 26

E

efekty dźwiękowe, 30
ekran
 aplikacji, 101
 dotykowy, 25, 26
 LCD, 26
 ładowania, 38
elementy
 grupujące, 39
 okna głównego, 60

F

formularz, 38, 78–82
 jakość, 79
 konstruowanie, 78
 wypełnianie, 78, 99

G

generowanie mowy, 88
gesty, 91
 komunikacyjne, 91
 konwersacyjne, 91
 służące do manipulacji, 91
 sterujące, 91
głośniki, 30
graficzne nakładki ekranowe, 38
graficzny interfejs użytkownika, GUI, 13, 36
 kategorie kontrolek, 37
 narzędzia prototypowania, 106
GUI, graphical user interface, 13, 36

H

HIG, human interface guidelines, 14
 historia rozwoju interfejsów, 11
 HUD, head-up display, 123, 124

I

I³, Immersion, Imagination, Interaction, 114
 IDE, Integrated Application Development, 106
 ikona, 69
 animowana, 72
 dźwiękowa, 72, 87
 ikony, 38
 informacje o programie, 39
 inspektor właściwości, 39
 interakcja, 115
 z komputerem, 33
 interfejs
 BCI, 17
 CLI, 12
 dla osób niepełnosprawnych, 16
 dźwiękowy, 86
 graficzny, 74
 Metro, 73
 multimedialny, 92
 multimodalny, 92
 neuralny, 18
 NUI, 92
 ocena, 105, 110
 projektowanie, 93
 prototypowanie, 105, 106
 sprzętowy, 7
 tekstowy, 34
 użytkownika, 7
 w systemie multimedialnym, 84

J

język
 naturalny, 86, 89, 99
 poleceń wierszowych, 34, 99
 SQL, 35
 VRML, 114
 XAML, 109

K

kafelki, 69, 72
 karty, *Patrz* zakładki
 kask HMD, 29, 115, 118

kategorie kontrolek, 37
 klawiatura, 21
 programowa, 22
 specjalna, 17
 kolory, 39, 103
 komputer, 10
 komunikaty, 39
 konfiguracja multimedialna komputera, 84
 konsola
 operatorska, 21
 systemowa, 11, 35
 kontrolery gier, 24
 kontrolki, 37
 grupowanie, 38
 nawigacja, 38
 wprowadzanie danych, 38
 wykonywanie akcji, 38
 wyświetlanie informacji, 38
 konwencje kolorów, 103
 kursor, 74

L

LCD, 29
 LED, light emitting diode, 29
 linia statusu, 59
 liniowe struktury menu, 40
 listy, 39, 101

Ł

łącza, 39

M

manipulacja
 bezpośrednia, 73, 75
 językowa, 76
 manipulator, 26
 kulkowy, 26
 Microsoft Kinect, 75
 mapa strony internetowej, 48
 mechanizm przeciągnij i upuść, 108
 menu, 39
 drzewiaste, 43, 45
 grafy, 43
 kontekstowe, 52
 kontekstowe pulpitu, 54
 liniowe, 41
 podręczne, 52
 pojedyncze, 41

- przestrzenne graficzne, 46
- sieci, 48
- stałe, 50
- struktury grafowe, 49
- systemy, 39, 40
- miniaturowy dżojstik, 26
- monitor, 28
- mowa ciągła, 89
- multimedia, 83
- mysz, 22, 26
 - typu 3D, 23

N

- narzędzia prototypowania GUI, 106
- nawigacja, 115
- NLS, oN-Line System, 12
- NUI, natural user interface, 92
- numer telefonu, 81

O

- ocena interfejsu, 110
- OCR, optical character recognition, 32
- okna, 39, 58
 - aktywacja, 59
 - dialogowe, 12, 39, 56, 61
 - główne, 61
 - modalne, 62
 - otwarcie, 59
 - przesuwanie zawartości, 60
 - ramka, 59
 - rodzaje, 61
 - z zakładkami, 68
 - zamknięcie, 59
 - zarządzanie, 63
 - zmiana rozmiaru, 60
- okruszki, 39
- okulary
 - AR Google Glass, 124
 - OHMD, 123
- OLED, 29
- opisy, 38

P

- pad, 24
- panele, 39
- paski
 - adresowe, 39
 - informacyjne, 38

- menu, 39
- narzędziowe, 39, 54, 59, 107
- postępu, 38
- statusu, 38, 59
- pióro
 - światłne, 26
 - wskazujące, 25
- płytką dotykową, 24, 26
- podpowiedzi, 38
- poła
 - danych specjalnych, 81
 - tekstowe, 39
 - wielokrotnego wyboru, 39
 - wyboru, 56
- polecenia
 - menu File, 44
 - systemów operacyjnych, 35
- pomoc dla użytkownika, 104
- poszukiwanie informacji, 48
- powiadomienia, 38
- prezentacja informacji wyjściowej, 100
- program GIMP, 67
- projektowanie
 - ekranów aplikacji, 101
 - interfejsu, 93
 - analiza użytkowników, 94
 - analiza zadań, 94
 - menu, 95
 - pomoc dla użytkownika, 104
 - prezentacja informacji, 100
 - struktura, 95
 - wizualne, 107, 109
- prototypowanie, 105
- przechowywanie danych multimedialnych, 11
- przetwarzanie wsadowe, 36
- przyciski radiowe, 39

R

- RAD, rapid application development, 106
- ramka, 39
- reprezentacje drzewa, 39
- rękawice wirtualnej rzeczywistości, 26, 115
- rodzaje
 - okien, 61
 - urządzeń sterujących, 26
- rozpoznawanie
 - głosu, 90
 - mowy, 19, 32, 86, 89, 90
 - pisma, 32
 - znaków ciągłych, 91

rozwijalne listy, 39
rozwój interfejsów, 11
rysik, 26
rzeczywistość
 rozszerzona, AR, 114, 122
 wirtualna, VR, 113

S

SaaS, Software as a Service, 11
SAR, spatial augmented reality, 123
schematy poszukiwania informacji, 48
sensory
 biometryczne, 26
 głębi, 26
skrót klawiaturowy
 Alt+Esc, 64
 Alt+Tab, 63
 Ctrl+Alt+Delete, 63
 Windows+Tab, 64
SQL, structured query language, 35
sterowniki nożne, 26
struktura
 drzewa, 43
 interfejsu, 95
styl
 Aero, 96
 Luna, 96
style interakcji, 33, 99
sygnały akustyczne, 86
symulatory lotu, 121
synteza mowy, 86
system
 komputerowy, 10
 NLS, 12
systemy
 informacji głosowej, 86, 88
 mowy niezależnej, 86
 mowy zależnej, 86
 multimedialne, 83
 przestrzennej rzeczywistości wzbogaconej,
 123
 z sygnałami akustycznymi, 86

Ś

środowisko
 3D, 115
 wirtualne, 115

T

tablet piórkowy, 25
tabliczka graficzna, 26
touchpad, 24
trójwymiarowa wizualizacja, 120
TTY, TeleTYpe writer, 11
TUI, text-based user interface, 12
typy urządzeń wejściowych, 20

U

UDK, Unreal Development Kit, 59
unifikacja interfejsu, 73
urządzenia
 interakcji, 19
 sterujące, 26
 śledzące
 postać użytkownika, 26
 wzrok, 26
 typu VR, 117
 wejściowe, 20
 wizualne, 28
 wyjściowe, 27
USB, universal serial bus, 31
usługi sieciowe, 11

V

VR, virtual reality, 113
 urządzenia, 117
 zastosowania, 118
VRD, virtual retinal display, 123

W

widżet, 37
wiersz polecenia, 35
WIMP, 13, 36, 37, 69
wizualizacja wyników zapytania, 47
wprowadzanie danych, 39
wskaźnik, 38
 kulkowy, 23
 wstażka, 39
wybór
 kolorów, 103
 stylu interakcji, 98
 układu okien, 65
 urządzeń interfejsu, 98
 z menu, 99

WYSIWYG, 106, 107

wyświetlacz, 28

typu HUD, 123

typu OLED, 29

Z

zakładki, 39, 67

arkuszy, 68

zanurzenie, immersion, 114

zarządzanie wieloma oknami, 63

zastosowania

drzewa, 44

manipulacji, 75

VR, 118

PROGRAM PARTNERSKI

GRUPY WYDAWNICZEJ HELION



1. ZAREJESTRUJ SIĘ
2. PREZENTUJ KSIĄŻKI
3. ZBIERAJ PROWIZJĘ

Zmień swoją stronę WWW
w działający bankomat!

Dowiedz się więcej i dołącz już dzisiaj!

<http://program-partnerski.helion.pl>

Interfejs użytkownika to wizytówka każdej aplikacji komputerowej i strony WWW, a także brama zapewniająca dostęp do ich funkcji. Nawet najlepsze, najbardziej wydajne i oferujące największe możliwości oprogramowanie ma niewielkie szanse na rynkowy sukces, jeśli jego interfejs będzie toporny, brzydki lub nieintuicyjny, a używanie go — męczące. Ludzie po prostu lubią korzystać z ładnie i funkcjonalnie zaprojektowanego oprogramowania i chcą to robić w jak najprostszy sposób, nie tracąc przy tym zbyt wiele czasu na naukę.

Budowanie przyjaznych interfejsów użytkownika to sztuka, którą zdecydowanie warto opanować, gdy tworzy się oprogramowanie komunikujące się z ludźmi. Pomoże Ci w tym książka *Podstawy projektowania interfejsów użytkownika*, wprowadzająca w tę rozbudowaną i interesującą dziedzinę wiedzy. Znajdziesz tu opis wybranych urządzeń wejścia-wyjścia oraz stosowanych obecnie elementów interfejsów, informacje na temat sposobów tworzenia projektów interfejsów i ich rozwoju, a także praktyczne przykłady zarówno dobrych, jak i złych rozwiązań w tej dziedzinie.

- Podstawowe informacje na temat UI
- Sposoby interakcji człowieka z komputerem
- Graficzne interfejsy użytkownika i ich elementy
- Rozwiązania multimedialne i multimodalne
- Zasady projektowania interfejsów użytkownika
- Możliwości analizy i oceny interfejsów
- Interfejsy stosowane w VR i AR

Dobrze programujesz? Zaczynj też dobrze projektować interfejsy użytkownika!

Helion

księgarnia internetowa



<http://helion.pl>

zamówienia telefoniczne



0 801 339900



0 601 339900

Helion SA
ul. Kościuszki 1c, 44-100 Gliwice
tel.: 32 230 98 63
e-mail: helion@helion.pl
<http://helion.pl>

Sprawdź najnowsze promocje:
• <http://helion.pl/promocje>
Książki najchętniej czytane:
• <http://helion.pl/bestsellery>
Zamów informacje o nowościach:
• <http://helion.pl/novosci>

sięgnij po **WIĘCEJ**



KOD KORZYŚCI

ISBN 978-83-283-3790-9



9 788328 337909

Informatyka w najlepszym wydaniu

cena: 37,00 zł