

Tomasz ROGALSKI

Politechnika Rzeszowska

Robert WIELGAT

Państwowa Wyższa Szkoła Zawodowa w Tarnowie

KONCEPCJA UKŁADU STEROWANIA SAMOŁOTEM LEKKIM ZA POMOCĄ POLECEŃ GŁOSOWYCH

W pracy zaprezentowano koncepcję systemu sterowania samolotem asystującego pilotowi. System różni się od tradycyjnie stosowanych tym, że sterowanie odbywa się za pomocą poleceń wydawanych głosem. W pracy przeanalizowano możliwości wykonania systemu sterowania z wbudowanym modułem automatycznego rozpoznawania mowy dla samolotów ogólnego przeznaczenia. Podjęto próbę zdefiniowania głównych reguł, którymi powinien kierować się projektant systemu. Zdefiniowano i omówiono główne poziomy funkcjonalności prezentowanego systemu sterowania. Szczegółowo omówiono zastosowanie wybranych komend do bezpośredniego sterowania lotem samolotu. Podano również podstawowe wymagania, jakie powinny spełniać komendy głosowe. Zawarto również krótki opis eksperymentalnego systemu sterowania stanowiącego podstawę rozwijania systemów głosowego sterowania samolotami ogólnego przeznaczenia.

Wstęp

Wciąż rosnący ruch małych samolotów ogólnego przeznaczenia (MSOP) jest ściśle związany z poszerzeniem grupy osób wykorzystujących samoloty tego typu jako „podniebne samochody.” Małe samoloty zaczęto używać jako środek lokalnego transportu na średnich dystansach [1]. Są one używane przez osoby niebędące zawodowymi pilotami, które często nie posiadają zbyt dużego doświadczenia w pilotowaniu samolotów i mają odmienne podejście do tego zagadnienia niż piloci zawodowi [5]. Spodziewają się również innych właściwości pilotażowych samolotu. Wydaje się więc celowe podjęcie prac konstrukcyjnych nad samolotem wyposażonym w system awioniczny przeznaczony dla tej grupy użytkowników. Takie systemy są przedmiotem prac badawczych w wielu ośrodkach na całym świecie. Systemy te mają uprościć proces sterowania. Zawierają wiele rozwiązań technicznych ułatwiających pilotowanie, nawigację i obsługę systemów pokładowych [8, 10].

Większość rozwiązań ułatwiających sterowanie dotyczy zmian właściwości pilotażowych samolotu, który jednak wciąż jest sterowany za pomocą drążka,

wolantu lub ministerownicy. Wszystkie te systemy angażują ręce pilota do sterowania. Niniejszy artykuł proponuje inną koncepcję systemu ułatwiającego latanie małym samolotem. Obecnie jest już technologicznie możliwe wykonanie systemu awioniki, który będzie rozpoznawał komendy głosowe wydawane przez pilota. Polecenia głosem mogłyby dotyczyć obsługi systemów pokładowych i parametrów lotu samolotu. Najbardziej pożądaną sytuacją byłoby, gdyby pilot mówił do systemu w sposób naturalny, jak do innego człowieka. Ale na takie rozwiązanie trzeba jeszcze poczekać.

Praca ma na celu zdefiniowanie, określenie zasad działania systemu sterowanego poleceniami głosowymi pilota. Omawia również problemy związane z interpretacją rozpoznanych poleceń. Autorzy pragną zwrócić uwagę na fizyczną możliwość realizacji systemu, który będzie wykorzystywał polecenia głosowe do sterowania systemami pokładowymi lub ich wybranymi funkcjami.

1. Funkcjonalność systemu

Zanim zostanie omówione zagadnienie sterowania głosem statku powietrznego najpierw trzeba odpowiedzieć na kilka fundamentalnych pytań:

1. Jakimi funkcjami samolotu można sterować za pomocą komend głosowych?
2. Czy pilot może głosowo sterować lotem samolotu?
3. Na jakim poziomie pilot może sterować samolotem?

Podczas typowego lotu pilot steruje bardzo wieloma funkcjami samolotu. Te funkcje posiadają różne poziomy ważności i związany z tym różny wpływ na bezpieczeństwo lotu. Można zatem wyróżnić trzy poziomy funkcjonalności, na których będzie pracować system sterowania głosowego (SSG). Różnią się one między sobą obsługiwanymi funkcjami oraz zastosowanymi procedurami rozpoznawania mowy.

Poziom I. Sterowanie głosowe obejmuje jedynie dodatkowe funkcje niewpływające bezpośrednio na bezpieczeństwo lotu. Są to: *włącz/wyłącz odmrózanie*, *włącz/wyłącz światła kołowania*, *włącz/wyłącz ogrzewanie rurki pitota*. Procedura rozpoznawania mowy może być w tym przypadku aplikacją rozpoznającą hasła z mowy ciągłej (ang. *word spotting*). Dopuszczalne jest popełnianie nielicznych błędów przez procedurę, z zagwarantowaną możliwością ich poprawienia.

Poziom II. Sterowanie głosowe obejmuje funkcje wpływające na stan lotu: *podwozie schowaj/wypuść*, *kłapy duże/male/schowaj*. Autopilot może być aktywowany i dezaktywowany, a zasadniczo żadne krytyczne funkcje sterujące nie mogą być obsługiwane. Oznacza to, że funkcje sterujące kierunkiem lotu, kątem wzniesienia, stabilizacją szybkości lotu nie mogą być uruchamiane głosowo, jeżeli wcześniej nie zostaną zaimplementowane odpowiednie zabezpieczenia. System musi zostać wyposażony w funkcje chroniące przed wprowadzaniem

nieprawidłowych wartości parametrów stabilizujących lot oraz przed zbyt dynamicznymi manewrami. Procedura rozpoznawania mowy na poziomie II może być, podobnie jak na poziomie I, aplikacją rozpoznającą hasła z mowy ciągłej, przy czym wymagania co do skuteczności rozpoznawania mowy są większe w porównaniu z poziomem I. Jednakże wciąż dopuszcza się popełnianie błędów przez system z możliwością ich korekty.

Poziom III. Na poziomie tym dostępne są funkcje stabilizujące wysokość lotu statku powietrznego, np. funkcje stabilizujące orientację przestrzenną. Ponadto powinny być zaimplementowane funkcje wyprowadzające samolot z niebezpiecznych i nieprawidłowych stanów lotu. Trzeci poziom sterowania głosowego wymaga niemal 100% skuteczności rozpoznawania mowy oraz bardzo niewielkich czasów rozpoznawania. System rozpoznawania mowy na poziomie tym powinien być systemem rozpoznawania połączonych słów lub systemem rozpoznawania mowy ciągłej.

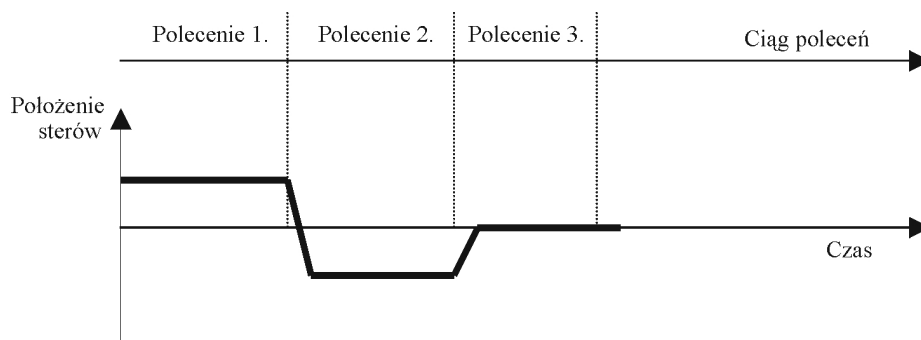
W celu zapewnienia bezpieczeństwa lotu wszystkie funkcje uruchamiane przez moduł rozpoznawania mowy muszą być dublowane przez tradycyjne urządzenia sterujące. Wszystkie rozpoznane komendy powinny być potwierdzane głosowo. Natomiast piloci powinni zostać przeszkoleni w zakresie korzystania z systemu rozpoznawania komend głosowych i sposobów radzenia sobie z potencjalnymi błędami rozpoznawania.

2. Sterowanie lotem

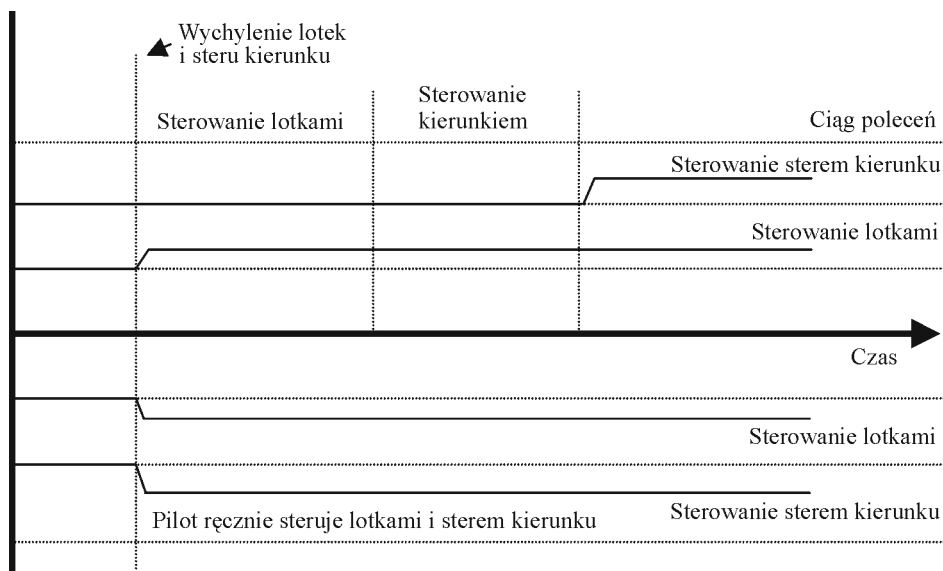
Na początku rozważań nad systemem głosowego sterowania samolotem należy zdefiniować sposób sterowania samolotem. Właściwości i charakter sterowania ręcznego sprawiają, że raczej niemożliwe będzie bezpośrednie sterowanie położeniem płaszczyzn sterowych za pomocą komend głosowych. Dzieje się tak, ponieważ:

- pilot wychyla płaszczyzny w sposób ciągły – w przypadku sterowania głosem będzie to sterowanie dyskretnie, nawet gdy komendy będą wydawane bardzo szybko (rys. 1.); dodatkowo nie jest możliwe natychmiastowe poprawienie niewłaściwego polecenia, jakiś czas samolot będzie sterowany źle, aż do momentu nadejścia nowej komendy; jak takie polecenia powinny być formułowane jako przyrosty czy bezwzględne wartości położenia sterów? Raczej trudno wyobrazić sobie sytuację, że przykładowo po podmuchu w skrzydło pilot będzie w stanie określić, o ile dokładnie wychylić lotkę, aby utrzymać stan lotu,
- w sterowaniu ma znaczenie nie tylko położenie końcowe płaszczyzn, lecz również sposób, w jaki zostało ono osiągnięte; ważne jest, czy płaszczyzny są wychylane dynamicznie czy wolno,

- pilot często steruje więcej niż jednym sterem w tym samym czasie, co nie byłoby możliwe w przypadku sterowania głosem (rys. 2.).



Rys. 1. Powierzchnie sterowe mogące zmieniać pozycje tylko dyskretnie

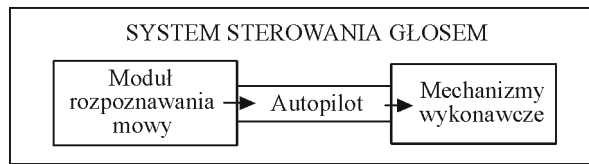


Rys. 2. Przykładowe przebiegi wychyleń powierzchni sterowych w klasycznym systemie sterowania i w systemie ze sterowaniem poleceniami głosowymi

Przedstawione spostrzeżenia prawdopodobnie przeważają na korzyść sterowania pośredniego. System sterowania będzie automatycznie utrzymywał parametry lotu zadawane przez pilota za pomocą poleceń głosowych. W praktycznej realizacji oznacza to integrację autopilota z modułem rozpoznawania głosu w jednym urządzeniu (rys. 3.). Autopilot będzie realizował zadania związane ze

stabilizacją parametrów lotu aktywowane przez moduł rozpoznawania i interpretacji poleceń głosowych.

Rys. 3. Połączenie systemu rozpoznawania komend głosowych z autopilotem



3. Możliwości realizacji

Nie ulega wątpliwości, że polecenia głosowe mogą być stosowane do sterowania samolotem jedynie w przypadku, gdy zastosowany sprzęt i algorytmy zapewnią szybkie i wiarygodne ich rozpoznawanie. System musi rozpoznać i zinterpretować polecenie w czasie krótszym niż pilot zdąży zauważyć opóźnienie. Biorąc pod uwagę zdolności percepcyjne człowieka, można zdefiniować czas reakcji systemu:

$$T_{SR} = \frac{1}{2} T_{HR} \quad (1)$$

gdzie: T_{SR} – czas, jaki system potrzebuje na rozpoznanie komendy i podjęcie właściwej akcji, T_{HR} – czas, jaki człowiek potrzebuje na wykrycie zdarzenia. Przy założeniu, że człowiek potrzebuje ok. $0,15 \div 0,2$ s na reakcję czas przetwarzania komendy przez system musi być krótszy niż $0,1$ s.

Algorytmy i platformy sprzętowe mogące znaleźć zastosowanie w opisywanym systemie są rozwijane i udoskonalane od wielu lat. Osiągnęły one poziom rozwoju umożliwiający ich zastosowanie w lotnictwie cywilnym. Telefonia cyfrowa od jakiegoś czasu oferuje głosowe wybieranie numeru lub obsługę centrów klienta. Podobne rozwiązania spotyka się w urządzeniach medycznych. Niestety wszystkie te rozwiązania działają według ściśle zdefiniowanych komend i nie są one w stanie rozpoznać polecenia znajdującego się w ciągu słów.

Powszechnie stosowane systemy nie rozpoznają poprawnie wszystkich poleceń, a skuteczność rozpoznawania zależy od wielu czynników: akcentu, intonacji, prędkości mówienia, głośności itp. Obecnie poziom rozpoznawania poleceń sięga 90° . W przypadku źle rozpoznawanych poleceń system nie może sterować żadną odpowiedzialną funkcją samolotu, w szczególności funkcjami związanymi bezpośrednio ze sterowaniem lotem. Jedynie funkcje nawigacyjne lub dodatkowe mogą być obsługiwane.

Biorąc pod uwagę ograniczenia systemów rozpoznających mowę, trzeba stwierdzić, że obecnie nie jest możliwe stworzenie kompletnego systemu sterowania głosowego w statkach powietrznych. Niemniej jednak można podejmo-

wać próby zmierzające do znalezienia optymalnego rozwiązania. Dalej przedstawiono rozważania dotyczące modułu rozpoznawania mowy w systemie sterowania samolotem.

W pierwszej kolejności należy zauważyć, że rozpoznawanie mowy w kokpicie samolotu dokonuje się w warunkach silnego zaszumienia sygnału oraz możliwych interferencji ze strony osoby towarzyszącej (drugiego pilota). Skuteczność rozpoznawania mowy jest dodatnio skorelowana ze stosunkiem sygnał-szum (SNR) dla sygnału mowy, dlatego każdy szum lub dźwięk pochodzący ze źródła innego niż rozpoznawany głos może spowodować drastyczne obniżenie skuteczności rozpoznawania mowy. Trudności te można częściowo zminimalizować, stosując specjalne mikrofony przyustne redukujące poziom szumów oraz nieczułe na dźwięki dochodzące z dalekiego pola. Jednakże wspomniane środki techniczne mogą okazać się niewystarczające. Dlatego pomocne mogą się okazać dodatkowe metody redukujące szum i hałas, np. filtry Wienera, macierze mikrofonowe, filtracja sygnału mowy. Oprócz tych środków wytłumienie kabiny pilota materiałami dźwiękochłonnymi może również przynieść dobre rezultaty.

W projektowaniu systemu sterowania głosowego samolotem należy również rozważyć rozmiar słownika oraz słowa podobne w brzmieniu (fonetycznie i akustycznie), lecz o różnym znaczeniu. Słowa te niejednokrotnie różniące się między sobą tylko jednym fonem sprawiają poważne trudności w rozpoznawaniu mowy. Ogólnie można powiedzieć, że im większy rozmiar słownika, tym mniejsza dokładność i dłuższy czas rozpoznawania. Z tego też względu zbiór rozpoznawanych słów powinien być możliwie jak najmniejszy. Można to częściowo osiągnąć przez dezaktywowanie rozpoznawania komend głosowych, których wydawanie w danej sytuacji nie ma sensu. Ponadto, jeżeli założyć się, że piloci korzystający z systemu zostaną wcześniej gruntownie przeszkoleni, rozmiar słownika można znacząco zmniejszyć. W takiej sytuacji nie trzeba wprowadzać nadmiernej ilości alternatywnych komend, które mają te same znaczenia, ale wypowiadane są na różne sposoby. Można również tak skonstruować słownik, aby zminimalizować liczbę kłopotliwych w rozpoznawaniu słów. Z drugiej strony, jeżeli system ma być całkowicie przyjazny dla użytkownika, rozmiar słownika musi być odpowiednio duży, aby zawrzeć w nim wszystkie alternatywy wypowiadania danej komendy głosowej. Implikuje to użycie bardzo szybkiego komputera pokładowego oraz odpowiednich algorytmów opartych na analizie dyskryminacyjnej oraz analizie składowych głównych (PCA).

Innym zagadnieniem, które należy rozważyć, jest zależność systemu od mówcy. Obecnie wiele programów rozpoznających mowę działa w sposób niezależny od mówcy. Jednakże procedury działające w trybie zależnym od mówcy pozwalają z reguły na osiągnięcie większej skuteczności niż procedury niezależne od mówcy. W przypadku opisywanego zastosowania stosunkowo łatwo można zaimplementować procedury zależne od mówcy, ponieważ pilot posiada

możliwość logowania się do systemu sterowania samolotem, jak również może nagrać wszystkie komendy przed pierwszym uruchomieniem systemu.

Istotnym parametrem, na który należy zwrócić uwagę w projektowaniu systemu sterowania głosowego, jest czas rozpoznawania. Rozpoznawanie powinno przebiegać w czasie rzeczywistym z maksymalnym opóźnieniem ok. 0,1 s. Krótki czas rozpoznawania posiada z reguły niekorzystny wpływ na skuteczność rozpoznawania.

4. Polecenia

Celem wyposażenia systemu awioniki samolotu lekkiego w układ rozpoznawania poleceń głosowych jest zwiększenie komfortu pracy pilota i zmniejszenie jego obciążenia. Nie może wystąpić sytuacja, gdy pilot musi szukać odpowiednich słów, aby prawidłowo zbudować rozkaz. W związku z tym forma wypowiedzianych poleceń musi być „przyjazna” dla użytkownika, co oznacza że:

- muszą to być standardowe frazy używane w lotnictwie – każde polecenie składa się z typowego połączenia słów używanych przez pilotów do określenia swoich oczekiwań co do parametrów lotu,
- muszą być maksymalnie skrócone – krótkie frazy jest łatwiej zapamiętać i można szybciej wymówić,
- akceptowane muszą być zarówno zmienne numeryczne, jak i lingwistyczne – „przyjemniej” jest powiedzieć „kurs *trochę* w prawo” niż formułować komendę bardzo precyzyjnie „kurs *trzy stopnie* w prawo”, chociaż druga forma jest niezbędna, gdy pilotowi zależy na dużej precyzji sterowania,
- zmienne lingwistyczne muszą mieć różne znaczenie zależnie od stanu lotu,
- różne frazy mogą mieć identyczne znaczenie – pilot nie może być zmuszony do użycia tylko jednej zmiennej lingwistycznej, np. wyrażenia „*lekko*”, „*trochę*” powinny mieć identyczne znaczenia,
- muszą być jednoznaczne – polecenia muszą się różnić w sposób zabezpieczający przed ich pomyleniem.

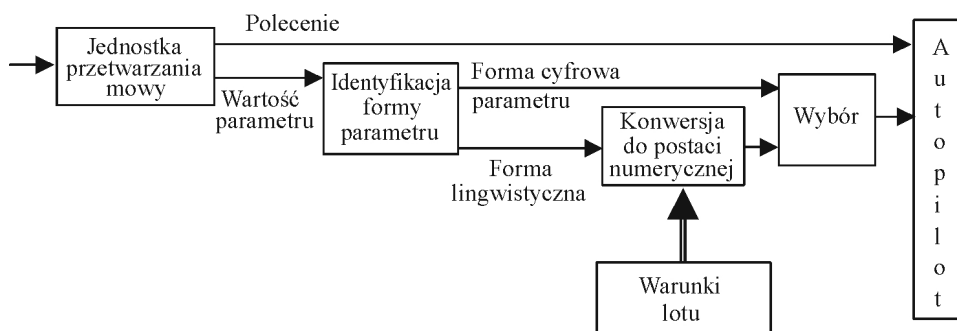
Zakłada się, że system rozpoznawania poleceń głosowych powinien umożliwić sterowanie systemami pokładowymi oraz samym lotem samolotu:

- polecenia *włącz/wyłącz*, np. *schowaj podwozie* lub *wysuń podwozie*,
- bezpośrednio definiujące wartość zadawanego parametru – *przechylenie dziesięć w prawo*,
- definiujące wartość sterowanego parametru przyrostowo – *przechyli w prawo dziesięć więcej*,
- specjalne polecenia na wypadek niebezpieczeństwa – polecenie zawsze stabilizujące lot horyzontalny *wyrównaj*.

Inna z kolei klasyfikacja poleceń uwzględnia sposób definiowania parametrów lotów:

- za pomocą dokładnej wartości numerycznej – dziesięć stopni, sto pięćdziesiąt węzłów,
- za pomocą zmiennej lingwistycznej – więcej, mniej.

Sam proces rozpoznawania i interpretacji poleceń składa się z następujących etapów (rys. 4.): moduł przetwarzania sygnałów dźwiękowych rozpoznaje polecenia i wartość parametrów. Następnie identyfikuje się, czy parametry są wartością liczbową czy zmienną lingwistyczną. Zmienne lingwistyczne są konwertowane na wartości liczbowe w zależności od stanu lotu samolotu. Następnie polecenie i parametry są konwertowane do postaci akceptowanej przez system sterowania samolotem i do niego transmitowane.



Rys. 4. Proces przetwarzania sygnału

W omawianym systemie wartość zmiennych lingwistycznych zależy od stanu lotu samolotu. Dlaczego? Przykładowo w sytuacji lotu z niezerowym kątem przechylenia pilot, chcąc zwiększyć przechylenie, wyda polecenie *przechyl więcej*. System musi zinterpretować, jaką wartość oznacza *więcej*. Ponieważ będzie to inna wartość dla już utrzymywanych małych i dużych kątów przechylenia, konwersja zmiennej lingwistycznej na wartość liczbową może odbywać się z wykorzystaniem kilku różnych metod. Można wykorzystać klasyczne zależności matematyczne:

$$v = f(x_1, x_2, \dots, x_n) \quad (2)$$

gdzie: v – wyliczany parametr, x_1, x_2, \dots, x_n – wielkości wpływające na wartość parametru wyznaczanego. Również metody logiki rozmytej lub sieci neuronowe mogą zostać wykorzystane do konwersji zmiennych lingwistycznych na wartości cyfrowe.

5. Moduł rozpoznawania mowy

Jak wcześniej wspomniano, moduł rozpoznawania mowy może być na poziomie I sterowania procedurą wyszukiwania haseł. Na najwyższym poziomie powinien znajdować się zorientowany zadaniowo system rozpoznawania połączonych słów lub system rozpoznawania mowy ciągłej. Wspólnymi etapami dla wszystkich tych systemów jest wstępne przetwarzanie sygnału oraz ekstrakcja cech.

Wstępne przetwarzanie sygnału

Do podwyższenia jakości procesu rozpoznawania mowy oraz racjonalnego wykorzystania zasobów systemu w systemach rozpoznawania mowy często stosuje się metody wstępnego przetwarzania sygnałów. Pierwszym urządzeniem w torze przetwarzania sygnału jest mikrofon. Z reguły są to mikrofony przyustne w zestawach nagłownych (słuchawkach lotniczych). Mikrofony te wraz z odpowiednimi urządzeniami mają możliwość tłumienia szumu i związanej z tym poprawy jakości rozpoznawanego sygnału mowy. Alternatywnym rozwiązaniem jest macierz mikrofonowa z zaprogramowaną procedurą formowania strumienia (ang. *beam-forming*). Technika formowania strumienia pozwala na wzmocnienie sygnału dźwiękowego (np. mowy pilota) z dowolnego kierunku oraz osłabienie sygnałów dochodzących z innych kierunków. Macierz mikrofonowa z zaimplementowaną techniką formowania strumienia może być wykonana w postaci samodzielnego urządzenia mikroprocesorowego, które nie wykorzystuje zasobów komputera pokładowego.

Sygnał otrzymany z mikrofonu lub macierzy mikrofonowej po ewentualnym wzmocnieniu powinien zostać przefiltrowany pasmowo do zakresu częstotliwości obejmującego widmo mowy. Filtracja pozwala na wyeliminowanie z sygnału składowych o niskich i wysokich częstotliwościach, leżących poza zakresem widma sygnału mowy, przez co składowe te nie będą negatywnie wpływać na jakość rozpoznawania mowy. Najwęższy zakres częstotliwości zapewniający jeszcze zrozumiałość mowy obejmuje od ok. 300 Hz do ok. 4 kHz. W taki sposób najczęściej filtrowany jest również sygnał akustyczny w kokpitach samolotów. Jednakże w przypadku systemów rozpoznających mowę zbyt duże zawężenie pasma sygnału może znacząco obniżyć skuteczność rozpoznawania. Optymalny zakres filtracji sygnału należy zatem dobrać eksperymentalnie.

Przefiltrowany sygnał akustyczny można poddać procesowi detekcji, który ma za zadanie wyodrębnić te fragmenty sygnału, które mogą być potencjalnie sygnałem mowy. Na dalszym etapie program analizuje jedynie te fragmenty, które zostały uznane za sygnał mowy, oszczędzając tym samym zasoby komputera i zmniejszając ryzyko błędnego rozpoznania sygnału.

Często spotykanym kryterium uznawania sygnału za sygnał mowy jest przekroczenie ustalonego wcześniej progu energii. Niektóre nowe techniki oparte są na detekcji sygnału z wykorzystaniem jego entropii.

Aby podnieść jakość sygnału mowy, można zastosować pewne techniki redukcji szumu, jak np. adaptacyjne filtry Wienera i Kalmana, metodę subtrakcji widma. Niemniej jednak należy mieć świadomość, że oprócz usuwania szumów wspomniane metody w pewnym stopniu zniekształcają również sygnał np. poprzez wprowadzenie tzw. szumu muzycznego. Najlepsza strategia odszumiania powinna być dobrana eksperymentalnie [4].

Ekstrakcja cech

Cechy sygnału mowy niezbędne do rozpoznawania mowy są ekstrahowane z wstępnie przetworzonego i odszumionego sygnału mowy. Obecnie najczęściej stosowaną metodą ekstrakcji cech sygnału mowy są parametry Mel-cepstralne (ang. *Mel-Frequency Cepstral Coefficients*, MFCC). Parametry MFCC oblicza się w następujących etapach:

- 1) podział sygnału na ramki, okienkowanie sygnału w ramkach za pomocą funkcji okna Hamminga (możliwe jest użycie również innych funkcji okna),
- 2) przeprowadzenie szybkiej transformaty Fouriera (FFT) na zokienkowanych sygnałach z ramek,
- 3) obliczenie zakumulowanych współczynników widmowych jako mocy FFT w tzw. krytycznych pasmach częstotliwości; pasma krytyczne zachodzą na siebie w 50%, jeżeli przyjęta jest melowa skala częstotliwości,
- 4) obliczenie logarytmu naturalnego zakumulowanych współczynników widmowych,
- 5) przeprowadzenie dyskretniej t ,
- 6) transformacji kosinusowej na zlogarytmowanych współczynnikach widmowych ($n = 0, 1, 2, \dots, q - 1$):

$$X(n) = c(n) \sum_{k=0}^{K-1} \ln(S_k) \cos\left(\frac{\pi(2k+1)n}{2K}\right) \quad (3)$$

$$c(0) = \sqrt{\frac{1}{K}}, \quad c(n) = \sqrt{\frac{2}{K}} \quad (4)$$

- 7) obliczenie pierwszej i drugiej pochodnej po czasie dla wektorów współczynników DCT, tzw. współczynników delta oraz delta-delta.

Rozszerzeniem współczynników MFCC są współczynniki HFCC (ang. *Human-Factor Cepstral Coefficients*, HFCC), które w pewnych warunkach są

bardziej odporne na szum niż cechy MFCC [6, 9, 12]. Różnica między parametrami MFCC i HFCC polega jedynie na obliczaniu szerokości krytycznych pasm częstotliwości. W przypadku parametrów HFCC szerokość pasma krytycznego oblicza się jako [3]:

$$ERB = 6,23f_c^2 + 93,39f_c + 28,52 \text{ Hz} \quad (5)$$

Ponadto szerokość pasma krytycznego dla cech HFCC może być dodatkowo skalowana przez współczynnik większy od 1, przez co uzyskuje się lepsze rezultaty rozpoznawania.

Procedura wyszukiwania haseł

Wynikiem ekstrakcji cech jest ciąg wektorów cech reprezentujących komendę głosową. Procedura wyszukiwania haseł rozpoznaje komendę ze słownika poprzez odszukanie ciągu wektorów cech najlepiej reprezentujących rozpoznawaną komendę. Wyszukiwanie haseł bardzo często jest oparte na metodzie nieliniowej transformacji czasowej. Główną ideą metody jest znalezienie ścieżki optymalnej z minimalnym kosztem przejścia z lewego dolnego rogu do prawego górnego rogu tzw. macierzy odległości lokalnych. Pojedynczy element d_{mn} macierzy jest równy odległości między m -tym wektorem cech (np. HFCC, MFCC) rozpoznawanej wypowiedzi a n -tym wektorem cech wzorca odniesienia. Miarą odległości między wektorami cech jest często odległość Euklidesowa. Zakumulowaną odległość w każdym punkcie ścieżki optymalnej otrzymuje się na podstawie rekursywnej procedury. Przykładowa procedura rekursywna podana jest w równaniu:

$$g(i, j) = \begin{bmatrix} g(i-1, j) + d(i, j) \\ g(i-1, j-1) + d(i, j) \\ g(i, j-1) + d(i, j) \end{bmatrix} \quad (6)$$

W celu normalizacji otrzymanego wyniku zakumulowany koszt jest zwykle dzielony przez współczynnik D :

$$D = \sqrt{N_p^2 + N_w^2} \quad (7)$$

gdzie: N_p – liczba wektorów cech wzorca odniesienia, N_w – liczba wektorów cech rozpoznawanego słowa.

Nie ma potrzeby obliczania odległości zakumulowanej przy użyciu wszystkich odległości lokalnych, zatem ścieżka wzdłuż której oblicza się odległość

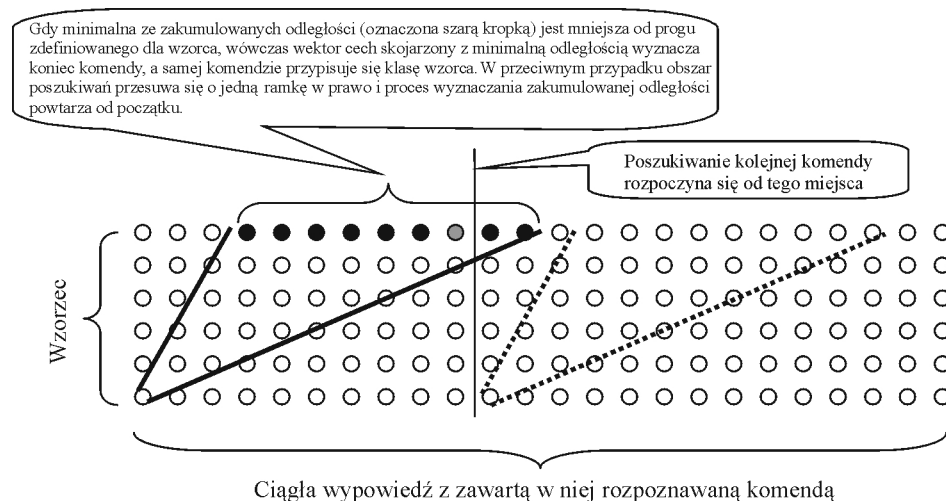
zakumulowaną leży wewnątrz obszaru w pobliżu przekątnej ograniczonego dwoma równoległymi liniami przesuniętymi o pewną liczbę (7):

$$Q = \text{round}(w \cdot \max(N_p, N_w)) \quad (8)$$

gdzie w – współczynnik przesunięcia.

Dokładny opis algorytmu DTW można znaleźć w pracy [11]. Po obliczeniu wszystkich końcowych zakumulowanych kosztów pomiędzy rozpoznawanym słowem oraz wszystkimi wzorcami za klasę słowa rozpoznawanego jest uznawana klasa wzorca dającego najmniejszy zakumulowany koszt.

Opisana procedura działa dobrze, jeżeli SNR sygnału mowy jest odpowiednio wysoki, a słowo, które ma być rozpoznawane, jest wykrywane za pomocą procedury detekcji sygnału mowy. Jednakże warunki takie są raczej niemożliwe do spełnienia w praktyce. Dlatego też bardziej efektywnym rozwiązaniem jest użycie metody DTW w procedurze wyszukiwania haseł. We wspomnianej procedurze zakłada się, że długość (w milisekundach) rozpoznawanego słowa jest przynajmniej $1/n$ długości wzorca oraz co najwyżej n razy większa od długości wzorca – zazwyczaj $n = 2$.



Rys 5. Znajdowanie komend za pomocą procedury wyszukiwania haseł

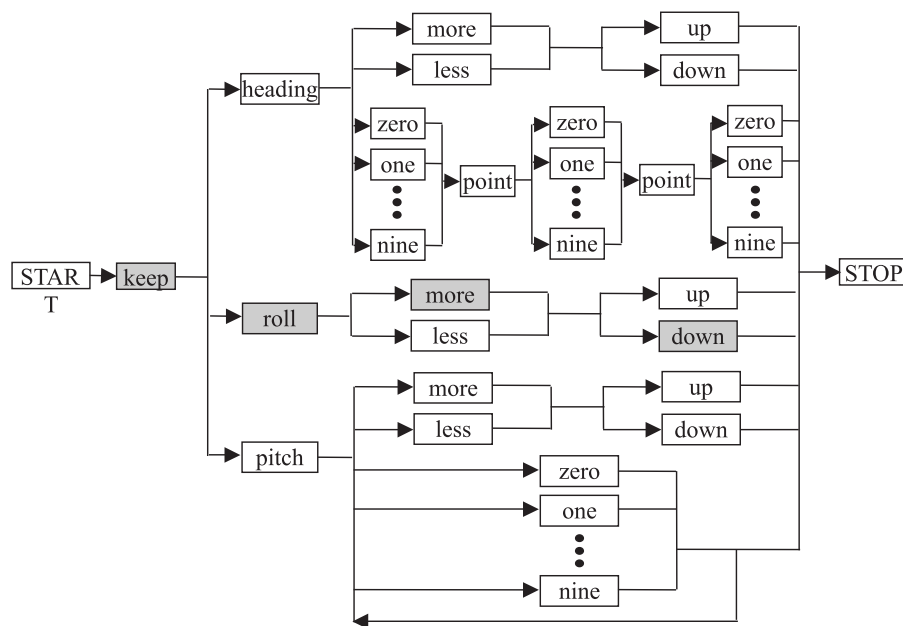
Założenia te powodują, że obszar poszukiwań ścieżki optymalnej przyjmuje kształt ograniczony ciągłymi liniami na rys. 5. Najniższy spośród zakumulowanych odległości na górnej granicy wskazuje koniec rozpoznawanej komendy, a ta odległość jest mniejsza niż próg dla danego wzorca. Klasa wzorca jest przypisywana do komendy. Poszukiwanie następnej komendy zaczyna się od następ-

nego wektora cech tuż za końcowym wektorem cech aktualnie rozpoznawanej komendy.

Aby procedury rozpoznawania mowy oparte na wyszukiwaniu haseł zadziały, muszą zostać zebrane wzorce wypowiedzianych komend. Wzorce te mogą być wbudowane w SRM, a także mogą być uzupełnione przez wzorce komend nagrane przez pilota korzystającego z systemu. Dobrą okazją do nagrywania wzorców komend dla danego pilota jest czas treningu na symulatorze lotu.

6. Rozpoznawanie połączonych słów

Rozpoznawanie połączonych słów jest z reguły wykonywane za pomocą niejawnych modeli Markowa (ang. *Hidden Markov Model*, HMM) opartych na modelach jednostek subleksykalnych, np. fonemów. Najpopularniejszym modelem Markowa fonemu jest model oparty na mieszaninie wielowymiarowych rozkładów Gaussa. Z reguły jest to model trójstanowy, nazywany często trójfonem. Procedurą trenującą modele HMM dla fonemu jest zwykle uczenie zintegrowane, będące odmianą procedury Bauma–Welcha. Rozpoznawanie odbywa się na podstawie algorytmu przekazywania żetonu, będącego zmodyfikowanym algorytmem Viterbiego [2].



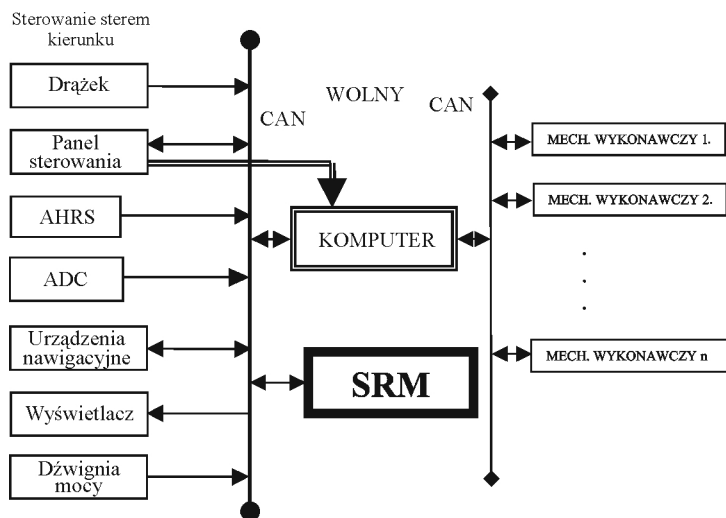
Rys. 6. Przykładowa sieć słów dla wybranej komendy systemu sterowania głosowego samolotem

Bardzo ważnym elementem procesu rozpoznawania połączonych słów jest zaprojektowanie odpowiedniej sieci słów, wewnątrz której program rozpoznający znajdzie najbardziej prawdopodobną hipotezę wypowiedzi. W projektowaniu sieci słów należy uwzględnić słowa o tym samym znaczeniu, lecz różniące się wymową (zapisem fonetycznym), oraz komendy alternatywne. Na rysunku 6. przedstawiono przykładową sieć słów służącą do sterowania głosowego pewnym elementem systemu awioniki. Sieć nie uwzględnia różnej wymowy słów, ale pewne komendy alternatywne. Gdy program znajdzie najbardziej prawdopodobną hipotezę wypowiedzi (zaznaczoną szarym kolorem na rys. 6.), wówczas komenda głosowa jest interpretowana tak, jak to przedstawiono na rys. 6.

Dokładny opis niejawnych modeli Markowa przekracza zakres prezentowanej pracy. Można go jednak znaleźć w literaturze [13].

7. Eksperymentalny system sterowania

Katedra Awioniki i Sterowania Politechniki Rzeszowskiej pracuje nad systemami sterowania dla samolotów lekkich już od wielu lat. Opracowano system sterowania SPS-1 [8, 10] składający się z szeregu niezależnie działających elementów, połączonych za pomocą magistrali cyfrowej CAN-2B [7]. Zastosowana w nim struktura sprzętowa oraz modułowość oprogramowania autopilota umożliwiają wymianę jego podzespołów w zależności od realizowanych zadań i implementację praw sterowania na innych platformach sprzętowych. Na przykład



Rys. 7. Ogólna struktura systemu sterowania z zaimplementowanym modulem rozpoznawania poleceń głosowych

na potrzeby projektu realizowanego wspólnie z Instytutem Lotnictwa system został zamontowany na pokładzie samolotu I-23 ze zmienionym modułem komputera pokładowego.

Struktura SPS-1 (rys. 7.) umożliwia połączenie z modułem rozpoznawania i interpretacji poleceń głosowych za pomocą magistrali. Do komputera pokładowego z tego modułu będą transmitowane polecenia sterujące lotem i systemami pokładowymi, powstałe wskutek działania algorytmów przetwarzania sygnału mowy pochodzącego od pilota – operatora systemu.

8. Podsumowanie

Systemy awioniki wspomagające pilota w różny sposób są obecnie coraz częściej spotykane. Tak więc system umożliwiający pilotowi głosowe sterowanie urządzeniami pokładowymi czy wybranymi funkcjami lotu samolotu z pewnością znajdzie użytkowników. Technicznie budowa takiego systemu zaczyna być możliwa. Pierwsze prototypy już się pojawiają. Pierwsze montowane seryjnie systemy będą to prawdopodobnie systemy poziomu pierwszego. Pozostałe, wymagające dużo skuteczniejszego procesu rozpoznawania i interpretacji poleceń głosowych, muszą być wciąż rozwijane i testowane.

Obecnie wydaje się, że prezentowane rozwiązanie może dać wiele korzyści, np. zwiększenie komfortu pracy pilota, który staje się bardziej osobą zarządzającą niż pilotem. Pilot, obsługując systemy pokładowe, nie będzie musiał odrywać wzroku od przestrzeni wokół samolotu. W przyszłości pilot równocześnie ze sterowaniem samolotem mógłby także wykonywać inne zadanie niezwiązane bezpośrednio z wykonywanym lotem.

Praca omawia wybrane problemy mogące pojawić się w systemie awioniki wyposażonym w układ rozpoznawania poleceń głosowych. Pokazuje możliwe drogi rozwoju systemów tego typu oraz ogólne zasady ich wykorzystania. Praca nie daje odpowiedzi na wszystkie możliwe pytania i wiele z nich z pewnością wciąż pozostaje bez odpowiedzi. Jednak w opinii autorów w przyszłości systemy umożliwiające głosowe sterowanie lotem samolotu na pewno pojawią się na pokładach samolotów ogólnego przeznaczenia.

Literatura

- [1] Kocks K.: Systems that permit everyone to fly, *Avionics Magazine*, March 2001, s. 16÷20.
- [2] Mohammed J.R.: A new robust adaptive beamformer for enhancing speech corrupted with colored noise, *Computer Systems and Applications, AICCSA, IEEE/ACS International Conference*, 31 March÷4 April 2008, s. 508÷515.
- [3] Moore B.C.J., Glasberg B., Baer T.: A model for the prediction of thresholds, Loudness, and Partial Loudness, *Journal of the Audio Engineering Society*, vol. 45, no 4, 1997, s. 224÷240.

- [4] Ouzounov A.: Robust features for speech detection – A comparative study, International Conference on Computer Systems and Technologies – CompSysTech 2005.
- [5] Rogalski T., Dolega B.: Algorithms improving flying qualities of general aviation aircraft, Aviation vol. X, no 2, 2006, s. 17÷22.
- [6] Sakoe H., Chiba S.: Dynamic programming algorithm optimization for spoken word recognition, IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-26, 1978, s. 43÷49.
- [7] Skowronski M.D., Harris J.G.: Exploiting independent filter band-width of human factor cepstral coefficients in automatic speech recognition, J. Acoust. Soc. Am., no 116(3) 2004, s. 1774÷1780.
- [8] Stock M.: Can aerospace, revision 1.6, Stock Flight Systems, Schutzenweg, Berg, Germany.
- [9] Tomczyk A.: Concept for simplified control of general aviation aircraft, World Aviation Conference, SAE/ACAA, Paper No 985551, Anaheim, CA 1998.
- [10] Tomczyk A.: Experimental fly-by-wire control system for general aviation aircraft, AIAA GNC Conference, Austin TX, AIAA Paper no 2003÷5776.
- [11] Wielgat R., Zieliński T.P., Woźniak T., Grabias S., Król D.: Automatic recognition of pathological phoneme production, Folia Phoniatica et Logopedica, vol. 60, no 6, 2008, s. 323÷331.
- [12] Wielgat R., Zieliński T.P., Świętojański P., Żołądź P., Król D., Woźniak T., Grabias S.: Comparison of HMM and DTW methods in automatic recognition of pathological phoneme pronunciation, INTERSPEECH 2007, Antwerp, Belgium 27÷31 August 2007.
- [13] Young S., Kershaw S., Odell J., Ollason D., Valtchev V., Woodland P.: The HTK Book (for HTK Version 3.0), <http://htk.eng.cam.ac.uk>.

A CONCEPT OF VOICE GUIDED GENERAL AVIATION AIRCRAFT

A b s t r a c t

The paper presents a conception of an airborne control system assisting the pilot, different than classical approaches present. There is a conception of voice commanded control system presented in this paper. The paper analyses possibilities of realization of aircraft control system with the speech recognition module for general aviation aircraft, at this moment. The author tries to define the main rules, which must be kept to project and operate such systems. The main functionality levels of such control system are defined and discussed too. The use of voice commands for the direct controlling the flight of the plane is also presented and analyzed in details. The paper discusses also types of voice commands pilot could use to control the plane. The main requirements voice commands must meet are defined and discussed. The last chapter contains the short description of the experimental control system, which can be the base for developing an experimental voice controlled general aviation aircraft.

Złożono w Oficynie Wydawniczej we wrześniu 2011 r.