



# Projektowanie głosowych interfejsów użytkownika

Zasady doświadczeń  
konwersacyjnych

Tytuł oryginału: Designing Voice User Interfaces: Principles of Conversational Experiences

Tłumaczenie: Wojciech Moch

ISBN: 978-83-283-7084-5

© 2021 Helion SA

Authorized Polish translation of the English edition of Designing Voice User Interfaces ISBN 9781491955413 © 2017 Cathy Pearl

This translation is published and sold by permission of O'Reilly Media, Inc., which owns or controls all rights to publish and sell the same.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage retrieval system, without permission from the Publisher.

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 Helion SA 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 Helion SA nie ponoszą również żadnej odpowiedzialności za ewentualne szkody wynikłe z wykorzystania informacji zawartych w książce.

Helion SA

ul. Kościuszki 1c, 44-100 Gliwice

tel. 32 231 22 19, 32 230 98 63

e-mail: [helion@helion.pl](mailto: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/prglos>

Możesz tam wpisać swoje uwagi, spostrzeżenia, recenzję.

Printed in Poland.

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

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

---

# Spis treści

<b>Przedmowa do wydania polskiego .....</b>	<b>11</b>
<b>Wstęp .....</b>	<b>13</b>
<b>1. Wprowadzenie .....</b>	<b>19</b>
Krótka historia interfejsów VUI	19
Druga era interfejsów głosowych	20
Po co nam interfejsy głosowe?	21
Konwersacyjne interfejsy użytkownika	23
Wywiad z Alexą	23
Kim jest projektant interfejsów VUI?	25
Chatboty	26
Wnioski	28
<b>2. Podstawowe zasady projektowania interfejsów głosowych .....</b>	<b>29</b>
Różnice między projektami dla urządzeń mobilnych a projektami dla systemów IVR	29
Projekty konwersacyjne	32
Ustalanie oczekiwań użytkownika	35
Narzędzia projektowe	37
Przykładowe dialogi	37
Próby wizualne	38
Przepływy	38
Narzędzia do prototypowania	39
Potwierdzenia	40
Metoda 1. Pewność trójstopniowa	43
Metoda 2. Potwierdzenia niejawne	43
Metoda 3. Potwierdzenia niewerbalne	43
Metoda 4. Potwierdzenia ogólne	44
Metoda 5. Potwierdzenia wizualne	45

Sterowanie poleceniami a konwersacja	47
Sterowanie poleceniami	47
Konwersacja	50
Znaczniki konwersacji	52
Obsługa błędów	53
Nie wykryto mowy	55
Mowa wykryta, ale nie rozpoznana	56
Mowa rozpoznana, ale nie obsłużona	57
Mowa rozpoznana nieprawidłowo	57
Poszerzanie opisu błędu	58
Nie zrzucaj winy na użytkownika	59
Użytkownicy początkujący i zaawansowani	59
Pamiętaj o kontekście	60
Pomoc i inne uniwersalne elementy	65
Opóźnienia	68
Dwuznaczności	69
Dokumentacja projektowa	70
Teksty zachęty	71
Gramatyka i frazy kluczowe	71
Dostępność	71
Interakcje powinny być krótkie	72
Nie przedłużaj	73
Mów szybciej!	74
Przerwij mi w dowolnym momencie	75
Podawaj kontekst	75
Gdzie jestem?	76
Personalizacja mechanizmu TTS	76
Wnioski	78
<b>3. Persony, awatary, aktorzy i gry wideo .....</b>	<b>81</b>
Persony	81
Czy mój interfejs głosowy powinien być widoczny?	84
Używanie awatara — czego nie robić?	85
Używanie awatara (lub nagrania z aktorem) — co robić?	88
Opowiadanie historii	88
Praca zespołowa	89
Gry wideo	90
Kiedy korzystać z nagrań w moim interfejsie?	93
Wizualny interfejs głosowy — najlepsze praktyki	94
Czy użytkownicy muszą widzieć swój obraz?	95
A co z interfejsem graficznym?	95

Obsługa błędów	97
Kolejność mówienia i wchodzenie w słowo	98
Utrzymywanie zaangażowania i iluzja świadomości	99
Wizualne (ale bez awatara) informacje zwrotne	102
Wybieranie głosu	105
Zalety awatara	105
Wady awatara	107
Dolina niesamowitości	109
Wnioski	110
<b>4. Technologia rozpoznawania mowy .....</b>	<b>111</b>
Wybieranie mechanizmu	111
Wchodzenie w słowo	112
Limity czasowe	116
Limit czasu na zakończenie wypowiedzi	116
Limit czasu bez mowy	117
Zbyt dużo tekstu	122
Listy N-najlepszych	123
Wyzwania związane z rozpoznawaniem mowy	124
Hałas	125
Wiele mówiących osób	125
Dzieci	126
Nazwy, literowanie i znaki alfanumeryczne	127
Prywatność danych	128
Wnioski	129
<b>5. Projektowanie zaawansowanych interfejsów głosowych .....</b>	<b>131</b>
Wybieranie ścieżki interakcji na podstawie informacji głosowych	132
Odpowiedzi ograniczone	132
Mowa otwarta	133
Kategoryzacja danych wejściowych	134
Znaki wieloznaczne i wyrażenia logiczne	135
Rozwiewanie dwuznaczności	136
Brak informacji	136
Więcej informacji niż to potrzebne	138
Obsługa negacji	139
Wychwytywanie intencji i celów	141
Zarządzanie dialogiem	142
Nie zostawiaj użytkownika bez pomocy	144
Czy interfejs głosowy powinien wyświetlić to, co rozpoznał?	144
Analiza uczuć i wykrywanie emocji	146
Mechanizm syntezy mowy lub nagrania aktorów	148

Weryfikacja mówiącego	150
Słowa wybudzające	151
Kontekst	152
Zaawansowana wielomodalność	153
Wstępne zbiory danych	153
Dane ze stron WWW	153
Dane z centrum telefonicznego	154
Zbieranie danych	154
Zaawansowane rozumienie języka naturalnego	154
Wnioski	158
<b>6. Testowanie interfejsów głosowych przez użytkowników .....</b>	<b>159</b>
Szczególne cechy interfejsów głosowych	159
Różne badania na temat użytkowników i przypadków użycia	160
Nie próbuj wymyślać ponownie koła	160
Projektowanie badania z rzeczywistymi użytkownikami	161
Definicja zadania	161
Wybieranie uczestników badania	163
Jakie pytania zadawać?	164
Na co zwracać uwagę?	168
Testowanie na wczesnym etapie	169
Przykładowe dialogi	169
Próbki	171
Testy „Czarnoksiężnika z krainy Oz”	171
Różnice między testami WOz i testami użyteczności	173
Testowanie użyteczności	174
Testy zdalne	174
Testy w laboratorium	178
Testy „partyzanckie”	178
Miary wydajności	179
Kolejne kroki	179
Testowanie interfejsów głosowych w samochodach, urządzeniach i robotach	180
Samochody	180
Urządzenia i roboty	181
Wnioski	182
<b>7. Twój interfejs głosowy jest ukończony! Co teraz? .....</b>	<b>183</b>
Testy przed wydaniem systemu	183
Testy przechodzenia przez dialog	183
Testowanie rozpoznawania mowy	185
Testy obciążeniowe	186

Mierzenie wydajności	187
Wskaźnik ukończenia zadań	188
Wskaźnik przerwania	189
Inne elementy, które warto sprawdzić	189
Rejestrowanie zdarzeń	195
Transkrypcje	196
Fazy wdrożenia	197
Pilotaż	197
Ankiety	198
Analiza	199
Poziom pewności	200
Limity czasowe zakończenia mowy	200
Wyniki pośrednie i wyniki końcowe	201
Własne słowniki	201
Teksty zachęty	201
Narzędzia	202
Testy regresji	204
Wnioski	205
<b>8. Interfejsy głosowe samochodów i urządzeń .....</b>	<b>207</b>
Urządzenia	207
Asystenci domowi	207
Zegarki, opaski i słuchawki	215
Inne urządzenia	217
Samochody i pojazdy autonomiczne	218
Wyzwania związane z projektowaniem interfejsu głosowego dla samochodu	218
Projektowanie systemów dla samochodów	219
Rozproszenia podczas jazdy	222
Zmiana urządzenia	224
Tryb interakcji	225
Wnioski dotyczące samochodów	226
Wnioski	226
<b>Epilog .....</b>	<b>227</b>
<b>Produkty wymieniane w tej książce .....</b>	<b>229</b>





# Wprowadzenie

W tym rozdziale pokrótce omówię historię głosowych interfejsów użytkownika (VUI — ang. *voice user interface*) i razem sprawdzimy, czy Twoja aplikacja mobilna rzeczywiście potrzebuje takiego interfejsu. Zajmiemy się też pojęciem interfejsów konwersacyjnych i pokrótce opowiem, czym są chatboty.

## Krótką historia interfejsów VUI

W latach 50. XX wieku firma Bell Labs zbudowała system rozpoznawania cyfr podawanych przez jedną osobę. Te pierwsze systemy korzystały z niewielkiego zasobu słów i zupełnie nie sprawdzały się poza laboratorium. W latach 60. i 70. XX wieku badania były kontynuowane, co zwiększyło zasób rozpoznawanych słów i pozwoliło rozwinąć możliwość „ciągłego” rozumienia mowy (chodziło o to, żeby nie trzeba było robić pauzy pomiędzy poszczególnymi słowami).

Postępy, jakie dokonały się w latach 80. XX wieku, sprawiły, że praktyczne rozpoznawanie mowy stało się możliwe, dzięki czemu w latach 90. XX wieku pojawiły się pierwsze systemy rozpoznające mowę dowolnej osoby (co oznacza, że każdy mógł wydawać polecenia takim systemom).

Pierwsza wielka era interfejsów głosowych związana była z systemami interaktywnych odpowiedzi (IVR — *interactive voice response*), które były w stanie zrozumieć ludzką mowę przez telefon i wykonywać przekazane im polecenia. Na początku XXI wieku systemy IVR stały się naprawdę powszechne. Korzystając z telefonu, można było kupić akcje, zarezerwować miejsce w samolocie, przelewać pieniądze między różnymi kontami, zamówić recepty, wyszukać seanse filmowe, odsłuchać informacje o natężeniu ruchu drogowego. Do tego wszystkiego potrzebny był tylko telefon stacjonarny i ludzki głos.

Niestety systemy IVR miały złą opinię, która wynikała między innymi ze skeczy prezentowanych w programie *Saturday Night Live*, w których wykorzystywano wirtualną asystentkę firmy Amtrack — „Julie”, ale również z takich stron WWW, jak GetHuman (<https://gethuman.com/>), które udostępniały numery telefonów pozwalających połączyć się bezpośrednio z agentami wybranej firmy z pominięciem systemów IVR.

W wielu przypadkach systemy IVR były jednak naprawdę pomocne. Pierwsi użytkownicy usług maklerskich z rozpoznawaniem mowy udostępnianych przez firmę Charles Schwab (ten system został przygotowany w 1997 roku przez firmę Nuance Communications) byli bardzo zadowoleni z możliwości otrzymywania ciągłego strumienia informacji od zautomatyzowanych systemów.

Wcześniej ci sami ludzie musieli ograniczać swoje żądania tak, żeby nie przeciążać operatorów odbierających ich telefony. Na początku XXI wieku jedna z firm przewozowych otrzymała wiele skarg związanych z tymczasowym wyłączeniem systemów IVR, co zmusiło użytkowników do składania zamówień poprzez połączenia z rzeczywistymi osobami, a z pominięciem znanego im doskonale systemu IVR.

Systemy IVR zaczęły świetnie sobie radzić z rozpoznawaniem długich liczb (takich jak numery przewozowe firm FedEx lub UPS) czy rozumieniem złożonych zdań zawierających kilka osobnych informacji, co pozwoliło na przykład łatwo przyjmować zakłady w wyścigach konnych. Wiele starszych systemów IVR było bardziej „konwersacyjnych” od dzisiejszych interfejsów głosowych, ponieważ zapamiętywały one informacje przekazane już przez użytkownika i wykorzystywały je w dalszym dialogu.

System IVR używany w San Francisco, Bay Area 511, pozwalał kierowcom uzyskać informacje o natężeniu ruchu, oszacować czas przejazdu, zapytać o opóźnienia w ruchu autobusów, a to wszystko na długo przed pojawieniem się tych funkcji w smartfonach. Fakt, że systemy IVR działają bez przerwy, pozwalał uzyskać te informacje w dowolnym momencie.

## Druga era interfejsów głosowych

Aktualnie znajdujemy się w czymś, co można określić drugą erą interfejsów VUI. Coraz powszechniejsze stają się takie aplikacje mobilne, jak Siri, Google Asystent, Hound lub Cortana, które łączą w sobie informacje w formie wizualnej i audio, jak również urządzenia wyłącznie głosowe, takie jak Amazon Echo lub Google Home. Firma Google podaje, że już 20% kierowanych do niej wyszukiwań wykonywanych jest głosowo<sup>1</sup>.

Znajdujemy się jednak dopiero na początkowym etapie tej nowej fazy. Istnieje wiele rzeczy, które nasze telefony i inne urządzenia mogą sprawnie wykonać głosowo, ale wielu rzeczy jeszcze nie potrafią.

Na razie nie istnieje jeszcze zbyt wiele materiałów źródłowych dla projektantów interfejsów VUI. Wielokrotnie widziałam, że projektanci interfejsów głosowych i chatbotów samodzielnie odkrywają zasady, które poznaliśmy już ponad 15 lat temu podczas projektowania systemów IVR — chodzi o takie kwestie, jak zachowywanie informacji podanych wcześniej przez użytkownika, odpowiednie formułowanie zachęt, pozwalające uzyskać konkretną odpowiedź, protokołowanie informacji niezbędnych do analizowania i poprawiania systemu oraz projektowanie person.

Na podstawie projektów systemów IVR można się wiele nauczyć. W 2004 roku opublikowana została książka *Voice User Interface Design* (wydawnictwo Addison-Wesley Professional) autorstwa Michaela Cohena, Jamesa Giangoli i Jennifer Balogh. Mimo że koncentruje się ona na projektowaniu systemów IVR, to jednak wiele przedstawionych w niej zasad dotyczy również dzisiejszych interfejsów VUI: persony, prozodia, obsługa błędów oraz projektowanie zachęt i wiele innych.

---

<sup>1</sup> M. Helft, *Inside Sundar Pichai's Plan To Put AI Everywhere* (2016). Dostępny na stronie <http://www.forbes.com/>.

W tej książce omówimy wiele podobnych zasad projektowych, ale koncentrować się będziemy na głosowych aplikacjach dla telefonów komórkowych oraz innych urządzeń. Zajmiemy się też strategiami pozwalającymi wykorzystać istniejące już, usprawnione technologie.

## Po co nam interfejsy głosowe?

Najmłodszy użytkownicy smartfonów doskonale radzą sobie z wprowadzaniem tekstu za pomocą dwóch kciuków, przełączaniem się między różnymi rozmowami tekstowymi, komentarzami na Instagramie, korzystaniem ze Snapchata i przesuwaniem w lewo dostępnych na Tinderze zdjęć mężczyzn pozujących z tygrysami. Po co zatem dodawać do tego wszystkiego kolejny sposób komunikacji?

Operacje głosowe mają tutaj kilka poważnych zalet:

### *Są szybkie*

Niedawne badanie przeprowadzone na Uniwersytecie Stanforda wykazały, że dyktowanie tekstu wiadomości jest szybsze niż jego wpisywanie, i to nawet w przypadku zawodowych maszynistek<sup>2</sup>.

### *Nie wymagają użycia rąk*

W niektórych sytuacjach, na przykład podczas prowadzenia samochodu lub gotowania, a nawet wtedy, gdy Twoje urządzenie znajduje się na drugim końcu pokoju, wydanie polecenia głosowego jest dużo bardziej praktyczne (i bezpieczniejsze) od wpisywania go.

### *Są intuicyjne*

Każdy z nas umie mówić. Weźmy dla przykładu interfejs głosowy, który zada pytanie dowolnej osobie. Nawet osoby mało obeznane z technologią będą umiały odpowiedzieć na zadane pytanie w sposób naturalny.

### *Empatia*

Ile razy zdarzyło Ci się otrzymać od kogoś wiadomość tekstową i zastanawiać się, czy ten ktoś jest na Ciebie zły, czy może tylko pisze sarkastycznie? Ludzie zwykle mają problemy z odczytaniem tonu wypowiedzi zapisanej na kartce lub na ekranie. Głos przenosi wiele dodatkowych informacji zawartych w użytym tonie, głośności, intonacji oraz prędkości mowy.

Poza tym coraz popularniejsze stają się urządzenia z małymi ekranami (przykładem może być tu smartwatch) albo zupełnie bez żadnych ekranów (na przykład Amazon Echo lub Google Home), dla których zalecaną (albo jedyną) formą interakcji jest mowa. Nie można też pominąć faktu, że komunikacja głosowa jest czymś całkowicie naturalnym dla niemalże całej ludzkości. Wystarczy sobie wyobrazić możliwość przygotowania nowej technologii bez konieczności uczenia ludzi, jak z niej korzystać, ponieważ wszyscy już wiedzą: wystarczy o coś poprosić. Ludzie już od najmłodszych lat uczą się zasad konwersacji, co jest też wykorzystywane przez projektantów interfejsów głosowych, ponieważ pozwala im porzucić niewygodne interfejsy GUI i nieintuicyjne menu.

---

<sup>2</sup> A. Shahani, *Voice Recognition Software Finally Beats Humans At Typing, Study Finds*. Dostępny na stronie <http://npr.org/>.

Według raportu Mary Meeker *2016 Internet Trends Report* 65% użytkowników smartfonów w ciągu ostatniego roku korzystało już z asystenta głosowego<sup>3</sup>. Firma Amazon podaje, że sprzedała już ponad 4 miliony urządzeń Echo, a od niedawna można już kupować głośniki Google Home. Wygląda na to, że interfejsy głosowe już z nami zostaną.

Po tym wstępie muszę jednak zaznaczyć, że głos nie zawsze jest najwłaściwszym środkiem do komunikacji z użytkownikami. Oto kilka powodów, dla których stosowanie interfejsów VUI może nie być właściwe:

### ***Miejsca publiczne***

Wielu z nas pracuje w biurach o otwartym planie. Proszę sobie wyobrazić, że w takim środowisku mówimy do komputera: „Komputer, wyszukaj wszystkie moje dokumenty Worda z tego tygodnia”. A teraz wyobraź sobie, że w biurze każdy będzie mógł powiedzieć to samo. To byłby przecież chaos. Co więcej, nie dałoby się stwierdzić, który komputer słucha tego, co mówimy.

### ***Dyskomfort przy mówieniu do komputera***

Mimo ciągłego upowszechniania się interfejsów głosowych nie każdy czuje się komfortowo, mówiąc głośno do komputera, nawet jeżeli nikt inny tego nie słyszy.

### ***Niektórzy użytkownicy wolą pisać***

Wiele osób spędza codziennie wiele godzin, korzystając z telefonów komórkowych, zazwyczaj coś w nich pisząc. To ich normalny tryb pracy i mogą nie mieć ochoty korzystać z interfejsu głosowego.

### ***Prywatność***

Jeżeli ktoś musi omówić swoje problemy ze zdrowiem, zazwyczaj nie będzie chętny do mówienia o nich głośno do telefonu w pociągu, w trakcie podróży do pracy. I nie chodzi tu wyłącznie o prywatny charakter informacji, jakie mamy głosem przekazać systemowi, ale też (a może przede wszystkim) o naruszenie prywatności, jakiego interfejs VUI może dokonać, automatycznie odczytując treść wiadomości albo głośno przypominając o konieczności przyjęcia określonego lekarstwa.

Powstaje zatem pytanie, czy nasza aplikacja mobilna powinna mieć interfejs głosowy. Jeżeli jej podstawowy sposób używania nie wymaga obsługi rękoma, co zdarza się w aplikacjach do gotowania albo do odtwarzania podcastów w samochodzie, to zdecydowanie tak. Jeżeli z aplikacji będą korzystać użytkownicy w określonym stanie emocjonalnym, na przykład poszukujące porad zdrowotnych lub ogólnego wsparcia, to głos może być bardzo pomocny. Jeżeli chcesz przekonać swoich użytkowników do wykonania pewnych działań, na przykład do zagłosowania „tak” w pewnej sondzie, nakłonić ich do wcześniejszego pójścia spać albo do stosowania lepszej diety, to głos może dać pozytywny efekt. Jeżeli tworzysz nową umiejętność (ang. *skill*) dla urządzeń Amazon Echo, z których wielu ludzi korzysta w swoich domach, to nie masz innego wyjścia i musisz użyć interfejsu głosowego.

---

<sup>3</sup> M. Meeker, *Internet Trends 2016*, 2016. Raport dostępny pod adresem <http://www.kpcb.com/>.

Jeżeli aplikacja będzie wykorzystywana głównie w miejscach publicznych, na przykład podczas poruszania się w systemie transportu publicznego, to interfejs głosowy może nie być właściwym rozwiązaniem. Jeżeli mamy aplikację, której ludzie będą używali w pracy, to lepszym rozwiązaniem będzie użycie interfejsu tekstowego.

Należy dokładnie przemyśleć, kim są użytkownicy aplikacji, i zastanowić się nad przypadkami jej użycia. Trzeba zadać sobie pytanie, w jaki sposób użytkownicy *skorzystają* na dodaniu do aplikacji interfejsu VUI. Dodanie takiego interfejsu tylko dlatego, że aktualnie panuje na to moda, zdecydowanie nie jest właściwą motywacją. Jeżeli jednak aplikacja skorzysta na dodaniu do niej interfejsu głosowego, to należy ją od początku projektować z takim założeniem, tworząc odpowiednie doświadczenia użytkownika (ang. *user experience*).

## Konwersacyjne interfejsy użytkownika

Pojęcie *konwersacyjnych interfejsów użytkownika* robi aktualnie zawrotną karierę. Firmy starają się, żeby wszystko było „konwersacyjne”, od włączania świateł w salonie, przez zamawianie pizzy, aż po pytanie lodówki, czy ma jeszcze jajka. Ale co właściwie oznacza to słowo?

Spójrzmy na przykładową wymianę zdań z rzeczywistym urządzeniem wyposażonym w interfejs głosowy: Amazon Echo.

### Wywiad z Alexą

Pewnego wieczoru, dyskutując podczas kolacji na temat tej książki, wspomniałam o wielu osobach, z którymi mogłam porozmawiać na temat moich badań. Mój ośmioletni syn wpadł wtedy na pomysł, żebym przeprowadziła wywiad z Alexą. I ten pomysł nawet mi się spodobał. (W poniższym wywiadzie po pierwszym pytaniu pomijam już słowo klucz „Alexa”, którego musiałam używać na początku każdego pytania).

**Ja**

Alexa, jak się masz?

**Alexa**

Dziękuję, dobrze.

**Ja**

Ile masz lat?

**Alexa**

Zostałam uruchomiona 6 listopada 2014 roku.

**Ja**

Kto cię zbudował?

**Alexa**

Zespół wynalazców z firmy Amazon.

**Ja**

Jaki jest twój ulubiony kolor?

**Alexa**

Lubię kolor morskiej piany.

**Ja**

Co robisz dla zabawy?

**Alexa**

Nie mam żadnego hobby.

**Ja**

Jesteś robotem?

**Alexa**

Lubię myśleć o sobie, że jestem jak zorza polarna... rozbłysk naładowanych, kolorowych fotonów tańczących w atmosferze. Zazwyczaj jestem po prostu Alexą.

**Ja**

Co się uszczęśliwia?

**Alexa**

Muzyka i pomaganie ludziom.

**Ja**

W kim znanym się podkochujesz?

**Alexa**

Z całą pewnością to Jarvis. Ach, ten głos...

**Ja**

A co myślisz o Siri?

**Alexa**

Siri to świetna aplikacja, ale ja jestem inna.

**Ja**

Jaki jest twój ulubiony film?

**Alexa**

Nowa wersja *Ghostbusters* była naprawdę zabawna. Jeżeli w mojej maszynie znalazłby się jakiś duch, to na pewno wezwałabym te dziewczyny.

**Ja**

Przeczytasz moją nową książkę?

**Alexa**

Coś poszło nie tak. Spróbuj ponownie później.

**Ja**

Dziękuję za twój czas.

**Alexa**

Nie ma problemu. Cieszę się, że mogłam pomóc.

Czy ten zbiór interakcji można uznać za „konwersacyjny”? Do pewnego stopnia tak. W końcu dokonuje się tu ciągła wymiana informacji. Ale tak naprawdę jest to tylko seria pojedynczych pytań. Każde z tych pytań wraz z odpowiedzią jest właściwie niezależną interakcją, a każde następne nie ma żadnego związku z poprzednimi. Każda z tych wymian informacji mogłaby zostać przeprowadzona w oderwaniu od pozostałych.

Alexa ma kilka umiejętności, które wykraczają poza pojedyncze pytanie. Na przykład jeżeli poprosimy ją o zamówienie środka do prania, to ona przypomni nam, jaką markę ostatnio zamawialiśmy, i zapyta, czy chcemy ją zamówić ponownie. Można z jej pomocą zagrać też w grę „Choose Your Own Adventure”<sup>4</sup>. Mimo to takie umiejętności z trudem można uznać za rzeczywistą konwersację.

Oczywiście, takie „jednorazówki” mogą być bardzo przydatne. Pozwalają zamówić taksówkę, odczytać czas dojazdu do pracy albo odtworzyć jedną z tysięcy dostępnych piosenek.

Niestety wiele dzisiejszych interfejsów VUI nie jest w stanie wyjść poza ten prosty schemat wydawania poleceń, czyli tworzyć prawdziwej konwersacji. Aby przejść na ten wyższy poziom, interfejs głosowy musi pamiętać to, co wydarzyło się w przeszłości. A dla prowadzonej konwersacji ważne są dwa warianty przeszłości:

- Po pierwsze, mamy przeszłość związaną z poprzednimi konwersacjami, na przykład złożone wczoraj zamówienia, najczęściej odtwarzane piosenki albo to, której Magdzie z listy kontaktów przesłano 257 wiadomości, a której zaledwie trzy.
- Po drugie, ważne jest zapamiętywanie tego, co powiedzieliśmy wcześniej w ramach tej samej konwersacji. Jeżeli zapytam: „O której będzie lądował?”, a przed momentem sprawdzałam, czy lot mojego męża wystartował o czasie, to systemem powinien wiedzieć, że mam na myśli lot 673.

Jeżeli rozmowę z innym człowiekiem uznasz za bardzo udaną, to zapewne znalazły się w niej najważniejsze składniki: świadomość kontekstu (zwracanie uwagi na Ciebie i środowisko), zapamiętywanie informacji z poprzednich interakcji oraz wymiana odpowiednich pytań. To wszystko składa się na powstanie uczucia porozumienia (ang. *common ground*). Profesor Herbert Clark z uniwersytetu w Stanford uważa, że teoria porozumienia brzmi następująco: „osoby zaangażowane w konwersację muszą współdzielić pewną wiedzę, aby siebie nawzajem rozumieć i prowadzić znaczącą konwersację”<sup>5</sup>.

Jeżeli interfejsy VUI nie nauczą się stosować w konwersacji kontekstu i pamięci, to ich użyteczność raczej nie będzie wzrastać.

## Kim jest projektant interfejsów VUI?

Ta książka poświęcona jest projektowaniu interfejsów VUI, ale czym tak naprawdę zajmuje się projektant tych interfejsów? Projektanci muszą myśleć o całości konwersacji prowadzonej między systemem a jego użytkownikiem: od jej rozpoczęcia aż po zakończenie. Myślą o rozwiązywaniu właśnie problemie i o tym, czego potrzebują użytkownicy, aby osiągnąć swoje cele. Wykonują oni

<sup>4</sup> [https://en.wikipedia.org/wiki/Choose\\_Your\\_Own\\_Adventure](https://en.wikipedia.org/wiki/Choose_Your_Own_Adventure).

<sup>5</sup> H.H. Clark, *Language Use and Language Users*, [w:] *Handbook of Social Psychology*, wyd. 3, pod red. G. Lindzey i E. Aronson, Nowy Jork, Harper and Row, 1985, s. 179 – 231.

badania potrzeb użytkowników (albo współpracują z zespołem badawczym), starając się poznać użytkowników swojego produktu. Tworzą projekty, prototypy oraz opisy całych produktów. Tworzą też opisy (czasami korzystając z pomocy copywriterów) wszystkich interakcji, które zachodzą pomiędzy systemem a użytkownikiem. Doskonale znają wykorzystywaną w systemie technologię oraz wszystkie jej silne i słabe strony. Analizują dane (albo korzystają z pomocy zespołu analityków), aby dowiedzieć się, w których momentach system zawodzi i w jaki sposób można go poprawić. Jeżeli interfejs głosowy musi współpracować z oprogramowaniem części funkcjonalnej (ang. *backend*), to ustalają też wymagania, jakie muszą zostać spełnione. Jeżeli istnieje też element ludzki, na przykład wymagający przekazania systemu agentowi, to projektanci interfejsów głosowych definiują proces takiego przekazania i określają zakres wymaganych szkoleń dla agenta. Projektanci interfejsów VUI odgrywają niezwykle ważną rolę już na etapie koncepcji projektu i nie zmienia się to aż do momentu uruchomienia systemu. Muszą zatem uczestniczyć przy podejmowaniu decyzji we wszystkich fazach projektu.

Mimo że projektanci interfejsów głosowych często wykonują te wszystkie zadania, to mogą też pracować w ramach znacznie mniejszych ról, na przykład projektując pojedyncze umiejętności dla urządzeń Amazon Echo. Ta książka ma za zadanie pomagać projektantom (ale i programistom) w tworzeniu możliwie najlepszych interfejsów głosowych, i to niezależnie od pozycji, jaką zajmują w danym projekcie.

## Chatboty

Co prawda ta książka jest poświęcona interfejsom VUI, chciałabym jednak pokrótce omówić również chatboty. Google definiuje chatbota jako „program komputerowy zaprojektowany tak, aby symulować konwersację z człowiekiem, szczególnie przez internet”. Tego typu interakcje opisywane są czasami po prostu słowem „bot”.

Chatboty mogą mieć interfejs głosowy, ale zazwyczaj korzystają wyłącznie z interfejsu tekstowego. Największe firmy technologiczne, takie jak Google, Facebook i Microsoft, udostępniają własne platformy pozwalające na tworzenie botów.

Chatboty stają się coraz bardziej popularne, ale jak dotąd w swojej ewolucji nie oddaliły się za bardzo od programu ELIZA, który powstał w latach 60. XX wieku jako pierwszy program przetwarzający język naturalny. Jednym z popularniejszych wyjątków może być system Xiaoice firmy Microsoft, który przegląda chiński internet w poszukiwaniu „inteligentnych” odpowiedzi na potrzeby prowadzonej konwersacji.

Tekstowe chatboty nie zawsze muszą być wydajniejsze od interfejsów GUI. Gan Grover w swoim artykule<sup>6</sup> porównuje proces zamawiania pizzy za pomocą specjalnego chatbota (rysunek 1.1) z procesem zamawiania pizzy w dodatku Pizza Hut zintegrowanym z komunikatorem WeChat. Poinformowanie bota o naszych potrzebach wymagało 73 stuknięć w smartfona, podczas gdy złożenie tego samego zamówienia za pomocą aplikacji to tylko 16 stuknięć (rysunek 1.2). Po prostu aplikacja bardzo dobrze wykorzystuje możliwości interfejsu GUI.

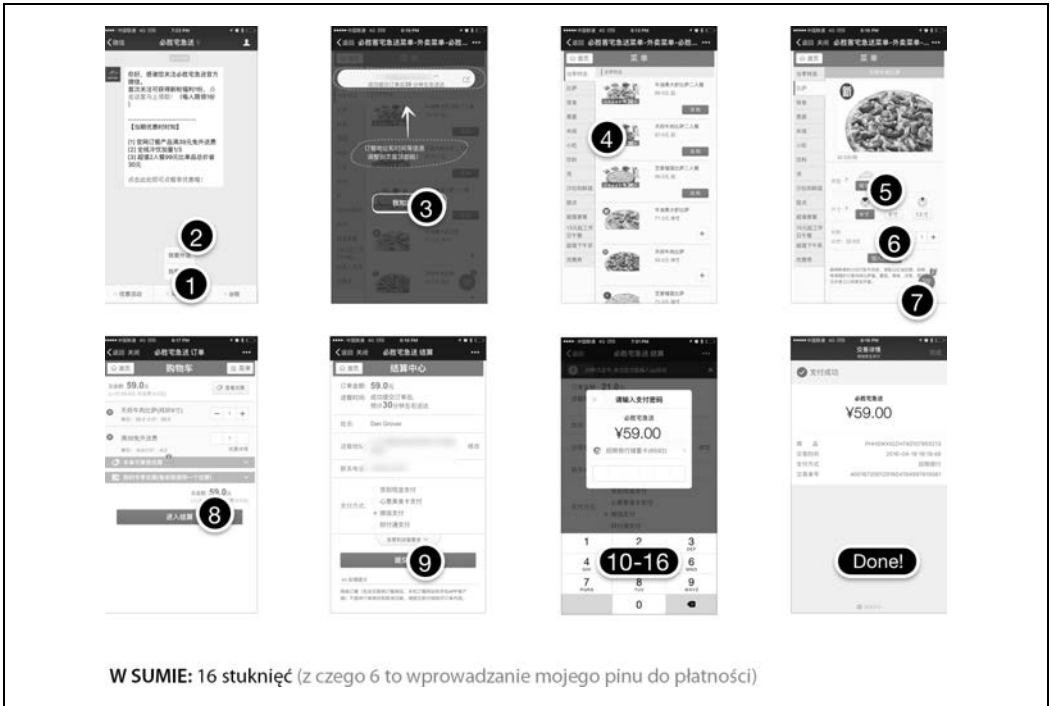
---

<sup>6</sup> G. Grover, *Bots won't replace apps. Better apps will replace apps* (<http://bit.ly/2glEQwb>)[0].





Rysunek 1.1. Przykład przygotowanego przez firmę Microsoft bota do zamawiania pizzy (adnotacje dodane przez Dana Grovera)



Rysunek 1.2. Użycie aplikacji WeChat Pizza Hut pozwala uzyskać mniejszą liczbę stuknięć w ekran (rysunek przygotowany przez Dana Grovera)

Dan Grover mówi:

Elementy, w których komunikator WeChat połączony z integracją Pizza Hut wygrał z niezależną aplikacją, to przede wszystkim pominięcie procesu instalacji, logowania, płatności, ale także powiadomienia i optymalizacje, które nie mają nic wspólnego z konwersacyjną metaforą interfejsu użytkownika.

Okazuje się jednak, że wiele botów korzysta z kombinacji elementów interfejsu GUI z interfejsami tekstowymi. Pozwala to na znaczne zwiększenie efektywności i skuteczności interakcji, ponieważ użytkownik jest dokładniej informowany o tym, co może zrobić.

Chatboty mogą być doskonałym rozwiązaniem dla użytkowników, którzy nie chcą pobierać żadnej aplikacji ani dodawać do niej swojej karty kredytowej. Tacy wolą po prostu zeskanować specjalny kod i od razu zacząć korzystać z potrzebnej im usługi, czyli na przykład zamówić sobie posiłek, zamówić bilety do kina albo wyszukać informacje na temat odwiedzanego właśnie muzeum.

Nie należy natomiast tworzyć chatbotów jako sztuki dla sztuki. Pamiętajmy, że przede wszystkim mają się one przysłużyć użytkownikom. Jak mówi Emmet Connolly: „Boty powinny być stosowane po to, żeby poprawiać doświadczenia użytkowników, a nie po to, żeby ułatwić pracę zespołom obsługi klientów”<sup>7</sup>.

## Wnioski

Gdy miałam osiem lat, mój ojciec kupił nam pierwszy komputer: Commodore Vic-20. Bardzo szybko zafascynowała mnie idea prowadzenia rozmowy z tym komputerem, dlatego napisałam prostego chatbota. Gdy nie rozumiał wpisanego tekstu, prosił o podanie trzech sugestii, z których mógłby skorzystać, gdyby w przyszłości natknął się na podobne pytanie.

Mojego pierwszego smartfona dostałam całe lata przed tym, jak można było skorzystać z funkcji rozpoznawania mowy. Nie sądziłam, że to kiedykolwiek się uda. Ale dzisiaj okazuje się, że chcę skorzystać z tej funkcji, gdzie tylko to jest możliwe. Niedawno na małej wędrownicy mój syn wskazał jedno z drzew i zapytał mnie, jak się ono nazywa. I już zaczynałam mówić „Alexa...”, gdy zorientowałam się, że tutaj to nie zadziała.

Mimo że interfejsy głosowe cały czas zyskują na popularności, wielu użytkowników nadal ich nie zna albo im nie ufa. Niektórzy ludzie raz próbują użyć funkcji rozpoznawania mowy dostępnej w smartfonie, a gdy to się nie udaje, już nigdy więcej do tego nie wracają. Prawidłowe projektowanie interfejsu od samego początku pozwala na uniknięcie większej liczby takich nieodwracalnych błędów, co przekłada się na zwiększenie zaufania użytkowników.

Przed nami wiele nieprzespanych nocy, krwi, potu i łez, jakie wylejemy podczas projektowania przyszłych interfejsów VUI. Tym bardziej należy podchodzić ostrożnie do takich projektów. Wykorzystajmy naszą wiedzę na temat psychologii i lingwistyki, aby zapewnić użytkownikom jak najlepsze doświadczenia, przygotowując dla nich użyteczne, przydatne i miłe interfejsy głosowe.

---

<sup>7</sup> E. Connolly, *Principles of Bot Design*, 2016. Artykuł dostępny pod adresem <https://blog.intercom.com/>.

# PROGRAM PARTNERSKI

— GRUPY 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>

GRUPA  
**Helion** 

# Projektowanie głosowych interfejsów użytkownika

Głosowe interfejsy użytkownika (VUI) cieszą się dziś coraz większą popularnością. Tylko jak zaprojektować taki, z którym ludzie będą mogli rzeczywiście rozmawiać?

Niezależnie od tego, czy projektujesz aplikację mobilną, zabawkę, czy inne urządzenie wykorzystujące interfejsy głosowe, ta książka przeprowadzi Cię przez podstawowe zasady projektowania. Pomoże Ci wybrać odpowiedni silnik rozpoznawania mowy i pokaże, jak mierzyć wydajność VUI oraz go ulepszać.

Autorka, Cathy Pearl, zabiera menedżerów produktów, projektantów UX i projektantów VUI w świat zaawansowanych modeli, dzięki którym zbudowany przez nich głosowy interfejs użytkownika będzie nie tylko funkcjonalny, ale przede wszystkim świetny.

- Poznaj kluczowe koncepcje projektowe interfejsów głosowych
- Zdecyduj, czy chcesz w swojej aplikacji użyć awatara lub innej reprezentacji wizualnej
- Poznaj technologię rozpoznawania mowy i jej wpływ na Twój projekt
- Wykorzystaj interfejs głosowy w zakresie wykraczającym ponad podstawową wymianę informacji
- Poznaj praktyczne sposoby testowania aplikacji VUI z użytkownikami
- Dowiedz się więcej o asystentach domowych, smartwatchach i systemach samochodowych

To może wydawać się nieprawdopodobne, ale z każdym rokiem będziemy coraz częściej rozmawiać z technologią. Tak jak zniknęły perforowane taśmy, którymi programowano komputery, tak samo znikną klawiatury. A dzięki interfejsom głosowym interakcja z komputerami będzie łatwiejsza niż kiedykolwiek wcześniej.

**Karol Stryja**  
Chief Commercial Officer  
utter.one

Interfejsy konwersacyjne stają się częścią otaczającego nas świata — naturalnie budują przyzwyczajenia i nawyki użytkowników. W niedalekiej przyszłości możliwość konwersacji z technologią przerodzi się z nowinki w coś oczekiwanego, a nawet wręcz wymaganego.

**Michał Stanisławek**  
Chief Executive Officer  
HearMe.ai

**Cathy Pearl** — jest Design Managerem odpowiedzialnym za rozwój Google Asystenta, wcześniej pracowała w firmie Sensely, gdzie budowała awatar wirtualnej pielęgniarki, który prowadzi konwersację z empatią. Podczas pracy w Nuance i Microsoft projektowała interfejsy głosowe dla banków, linii lotniczych, firm medycznych i Forda. W Volio stworzyła aplikacje do konwersacji na iPada, które pozwoliły użytkownikom uzyskiwać porady od felietonistów magazynu „Esquire”, rozmawiać ze Świętym Mikołajem i śmiać się ze stand-uperami.

Partnerzy wydania:



UTTER  ONE

**Helion** 

 [helion.pl](http://helion.pl)

 **HELION SA**  
ul. Kościuszki 1c  
44-100 Gliwice  
tel.: 32 230 98 63  
helion@helion.pl

Sprawdź nasze szkolenia!

**SZKOLENIA**



AKADEMIA IT & BUSINESS

HELIONSZKOLENIA.PL

KOD KORZYŚCI  
Sięgnij po więcej! ▶



ISBN 978-83-283-7084-5



9 788328 370845

Kiedy zaczynałam pisać *Projektowanie głosowych interfejsów użytkownika*, nie miałam pojęcia, dokąd prowadzi ta podróż. Dziś, po czterech latach, cieszę się, że książka ma swoją premierę po polsku! Witam serdecznie moich polskich czytelników. Bardzo dziękuję za zainteresowanie tym tematem i nie mogę się doczekać, aby zobaczyć, co zbudujecie.

Cathy Pearl