

**Grażyna Stachyra**<http://orcid.org/0000-0002-6396-112X>

Uniwersytet Marii Curie-Skłodowskiej w Lublinie

grazyna.stachyra@mail.umcs.pl

DOI: 10.35765/pk.2023.4203.14

## Głos jako element tożsamości w hybrydycznej rzeczywistości relacji człowiek – asystent głosowy

### STRESZCZENIE

Artykuł podejmuje problem asystentów głosowych, których algorytmy syntezy mowy imitują głos naturalny. Liczne badania dowodzą, że użytkownicy tych urządzeń oraz aplikacji traktują je jako istoty społeczne. Takie zachowanie jest efektem hybrydycznego charakteru rzeczywistości komunikacyjnej, w której cechy głosu naturalnego wywołujące określone reakcje są transponowane na głos przetworzony technologicznie. Diagnoza tej problematyki obejmująca aspekty specyfiki mowy syntetycznej oraz modelowania walorów emocjonalnych w głosie generowanym pozwala wnioskować, że człowiek intuicyjnie nadaje głosowi asystenta te cechy, które naturalny głos sygnalizuje w kontaktach międzyludzkich. Ponieważ ze społecznego punktu widzenia głos jest nośnikiem tożsamości mówiącego (czyli cech fizycznych, psychicznych i społecznych), artykuł wskazuje, dlaczego możliwość klonowania głosu burzy kulturowy porządek postrzegania głosu jako elementu tożsamości konkretnej osoby oraz dlaczego zakłóca świadomość odrębności własnego i cudzego głosu.

**SŁOWA KLUCZE:** asystent głosowy, głos, hybrydyczna rzeczywistość, komunikacja, sztuczna inteligencja, tożsamość

### ABSTRACT

Voice as an Identity Element in the Hybrid Reality of the Human-voice Assistant Relationship

The article addresses the problem of voice assistants, whose speech synthesis algorithms mimic the natural human voice. Numerous studies show that users of such devices and applications treat them as social beings. This behaviour results from the hybrid nature of the communication reality, in which the characteristics of the natural voice that evoke specific reactions are transposed onto a technologically processed voice. Diagnosis of this issue, which includes aspects of the peculiarities of synthetic speech and the modelling of emotional qualities in the generated voice, allows us to conclude that humans intuitively

**Sugerowane cytowanie:** Stachyra, G. (2023). Głos jako element tożsamości w hybrydycznej rzeczywistości relacji człowiek – asystent głosowy. © *Perspektywy Kultury*, 3(42), ss. 179–190. DOI: 10.35765/pk.2023.4203.14

Nadesłano: 24.03.2023

Zaakceptowano: 30.08.2023

give the voice of an assistant those qualities that the natural voice signals in interpersonal contacts. Since, from a social point of view, the voice is a carrier of the speaker's identity (i.e., physical, psychological, and social characteristics), the article shows why the possibility of cloning the voice shatters the cultural order of perceiving the voice as an element of a particular person's identity, and why it disrupts the awareness of the distinctiveness of one's own and another's voice.

**KEYWORDS:** voice assistant, voice, hybrid reality, communication, artificial intelligence, identity

### Hybrydyczna rzeczywistość komunikacji głosowej

Technologie sztucznej inteligencji obejmują różne funkcjonalności. Jedną z nich stanowią tzw. asystenci, komunikujący się głosem powstałym w wyniku działań algorytmów syntezy mowy. Najnowsze narzędzie sztucznej inteligencji opracowane w styczniu 2023 r. – VALL-E, wiernie imituje ludzki głos w całej jego naturalności na bazie 3-sekundowej próbki (Wang i in., 2003). To rewolucja w postrzeganiu głosu jako elementu tożsamości człowieka. Skoro bowiem silniki przetwarzania mowy AI potencjalnie mogą odtwarzać cechy dowolnego głosu, a badania (Ostrowski, Fu, Zygoras, Park i Breazeal, 2022) wykazały, że użytkownicy przypisują asystentom ludzkie cechy – tożsamość budowana na bazie głosu przestaje być dystynkcją człowieka. Powstaje hybrydyczna sfera komunikacji, w której człowiek wprawdzie uświadamia sobie, że rozmawia z robotem, ale słysząc sklonowany głos naturalny, intuicyjnie nadaje mu rys osoby. Warto zatem podjąć próbę zinterpretowania walorów głosu jako elementu tożsamości człowieka w kontekście takich kwestii, jak:

- wzrastająca rola komunikacyjna asystentów głosowych AI;
- transponowanie przez użytkowników cech osobowych na asystenta;
- świadomość głosowej odrębności.

Powiązanie głosu z tożsamością człowieka stanowi kulturowy ryt komunikacji międzyludzkiej. Także w komunikacji człowiek – maszyna głos ma istotne znaczenie. Okazuje się bowiem, że ludzie nie tylko traktują maszyny jako istoty społeczne (Nass, Steuer i Tauber, 1994; Guzman i Lewis, 2019), ale przypisują im różne cechy osobowości właśnie na bazie generowanego przez nie głosu (Nass i Lee, 2001). To głos odgrywa decydującą rolę wśród cech personifikujących technologię (Wagner i Schramm-Klein, 2019). Ponieważ brzmienie głosu jest podstawą precyzowania cech interlokutorów, indywidualizowanie głosu syntetycznego poprzez zmiany intonacji, tempa mówienia czy natężenia dźwięku – wpływa na wysoką

wiarygodność relacji człowiek – maszyna (Cambre i Chinmay, 2019). Tym samym głosowa komunikacja człowieka z asystentem nie jest już tylko technologicznie zapośredniczonym kontaktem, ale relacją, w której głos sztucznej inteligencji jest nośnikiem konkretnych cech osoby.

Skoro głos przestaje być jednostkowym i niepowtarzalnym składnikiem tożsamości człowieka, komunikacja wkracza w rzeczywistość hybrydyczną rozmywającą granicę pomiędzy dźwiękiem emitowanym przez człowieka a tym tworzonym przez maszynę. Jak twierdzi Pierpaolo Donati (2021), wskutek trwałych interakcji z technologiami „dotychczasowe sposoby pojmowania siebie i relacji zgodnie z zasadami tzw. analogowej rzeczywistości mieszają się z regułami rzeczywistości wirtualnej” (s. 218–220). Podobnie jak AI jest utożsamiana z kodem cyfrowym, ludzie są identyfikowani choćby ze swoimi wirtualnymi profilami na Facebooku. Algorytmy AI oficjalnie występują jako dziennikarze-roboty, a ich dokonania porównywane są z twórczością ludzi w dziennikarskich konkursach pisarskich (Smith, 2015, za: Guzman i Lewis, 2019). Ludzie postrzegają asystentów głosowych jako byty pomiędzy człowiekiem a przedmiotem (Chin i Robinson, 2020), zaś granica między tym, co ludzkie, a tym, co sztuczne, staje się coraz bardziej płynna (Turner 2019, za: Schröder, 2021).

## Wzrastająca rola komunikacyjna asystentów głosowych AI

Termin sztuczna inteligencja (AI, *Artificial Intelligence*) użyty po raz pierwszy przez Johna McCarthy’ego w 1956 r. został zdefiniowany jako inżynieria tworzenia inteligentnych maszyn (McCarthy, 2007). Rafał Garbacz, współzałożyciel start-upów UtterOne i HearMe.ai<sup>1</sup> zajmujących się technologiami głosowymi, podkreśla złożoność algorytmów AI:

AI nie jest jednym mechanizmem, ale zbiorem oddzielnych technologii odpowiedzialnych za różne zastosowania, które wzajemnie się łączą i przenikają. Asystent głosowy łączy w sobie liczne mechanizmy, jak: Text-To-Speech, Speech-To-Text, NLP, NLU, rozumienie, baza wiedzy, komponenty odpowiedzialne za *smart home* i wiele innych. To, co mówimy obecnie zamieniane jest na tekst. Silniki rozumienia NLU oraz NLP przetwarzają nasze intencje, zadają problem do zbioru danych czyli modelu, który tworzy z dostępnych danych odpowiedź, potem generowany jest

---

1 Start-up stworzył platformę, która umożliwia konwertowanie internetowych treści pisanych na podcast (treść audio) lub tzw. *flash briefing* (skrót wiadomości) realizowany głosami asystentów AI oraz automatyczne duplikowanie ich w dowolnych mediach społecznościowych użytkownika. Cytowane wypowiedzi pochodzą z wywiadu przeprowadzonego 13 lutego 2023 r. W tekście zostały oznaczone jako (RG, 2023)

tekst, który następnie jest syntezowany dowolnym głosem. To obecnie najpopularniejszy mechanizm, jednak technologia rozwija się szybko i powstają już prototypy, które nie wymagają sprrowadzania komunikacji do tekstu (RG, 2003).

Wirtualni asystenci posługujący się głosem sztucznej inteligencji (AI) wkroczyli na masowy rynek konsumencki po 2010 r. Najpierw Siri firmy Apple, a następnie między innymi Alexa firmy Amazon, Asystent Google'a i Cortana firmy Microsoft (Chin i Robinson, 2020). Zapotrzebowanie na asystentów wzrasta dynamicznie. Juniper Research (2018) wskazuje, że na świecie w użyciu są ponad trzy miliardy asystentów głosowych. Według niektórych szacunków ich liczba potroi się do końca 2023 r., osiągając 8 miliardów. Rozwój algorytmu pozwala AI coraz wierniej interpretować kształty fali głosowej rozrysowanej w tzw. spektrogramie, na którego podstawie tworzony jest zapis audio. Technologie rozpoznawania głosu oparte na uczeniu maszynowym (jak Amazon Alexa) doskonała się w identyfikowaniu biomarkerów wokalnych, czyli wzorców głosu charakterystycznych dla różnych chorób i schorzeń neurologicznych. W firmach handlowych asystenci głosowi AI badają parametry brzmieniowe klienta i sporządzają w czasie rzeczywistym profil jego zachowania uwzględniający nie tylko pozytywne lub negatywne nastawienie do zakupu, ale też na przykład skłonność do podejmowania ryzyka. Wirtualni asystenci prowadzą terapie dzieci z autyzmem, edukują i wchodzą w interakcje z podopiecznymi, wykorzystując ekspresyjność narracji głosowej i zdolność adaptacji walorów głosu do stanów afektywnych ludzi (Aeschlimann, Bleiker, Wechner i Gampe, 2020). Głos asystentów jest coraz subtelniej personalizowany, może naśladować głos naszych przyjaciół, bliskich, a nawet nas samych. Oznacza to, że zbiór cech głosowych stanowiących element tożsamości każdego człowieka wszedł „w posiadanie” technologii AI.

### Transponowanie przez użytkowników cech osobowych na asystenta

Valentina Pitardi i Hannah Marriott (2021) zauważają, że od czasu pojawienia się asystentów aktywowanych głosem rozmowy między ludźmi a maszynami skutkują wyższym poziomem zaangażowania interlokutorów, dzięki czemu wzrasta naturalność konwersacji, a nawet tworzy się paraspołeczna relacja pomiędzy rozmawiającymi. Głos jako ekspresyjne audialne medium komunikacji jest manifestacją dźwięku słyszanego przez odbiorcę i wokalizowanego przez nadawcę dzięki mechanizmom głosowym (Coeckelbergh, 2011). Mechanizmy te tradycyjnie odnosiły się

do wytwarzanych w ciele ludzkim lub zwierzęcym wibracji (fonowania głosu w krtani) oraz jego artykulacji w jamie ustnej, ale od czasu rozwoju mediów komputerowych i związanych z nimi technologii obejmują też maszynową syntezę dźwięku. Pojawia się zatem istotna kwestia migrowania tożsamościowych walorów głosu w przestrzeń technologii, ponieważ ze społecznego punktu widzenia głos jest sposobem projektowania przez mówiących swojej tożsamości – a więc cech fizycznych, psychicznych i społecznych – na świat (Kreiman, Vanlancker-Sidtis i Gerratt, 2003; McTear, Callejas i Griol, 2016).

Potwierdzają to badania na temat preferencji głosowych sygnalizujących płć asystenta AI<sup>2</sup>. Ankietowani precyzyjnie sklasyfikowali optymalny głos jako kobiecy, w wieku 20–30 lat – „ale nie w wieku średnim”, profesjonalny – „ale nie sztywny”, perswazyjny – „ale nie natrętny”, żartobliwy – „ale nie złośliwy” (Cambre i Chinmay, 2019, s. 13). Zwracają uwagę wskazane przez użytkowników niuanse odbioru głosu i wiążące się z nimi wyobrażenia konkretnych osób o konkretnej płci. Co ważne, w grupie młodych dorosłych wykazano różne wzorce mowy w stosunku do żeńskich i męskich głosów asystentów Apple oraz Siri odzwierciedlające wzorce z komunikacji międzyludzkiej. To sugerowałoby, że genderowe wzory w kontakcie człowieka z technologią wynikają nie tylko z utrwalonych doświadczeń komunikacyjnych, ale bywają też podświadome (Zellou i in., 2021). Stwierdzono również, że użytkownicy mieli tendencję do nadawania tożsamości płciowej nawet technologii celowo neutralnej pod względem płci. W badaniach z humanoidalnymi bezpłciowymi robotami badacz z Yale, Brian Scassellati, zauważył, że użytkownicy zwracali się do robotów per „on” lub „ona”, mimo że sami badacze stosowali określenie „to” (Chin i Robinson, 2020). W ten sposób głosowo sygnalizowana płć stawała się istotnym elementem osobowości asystenta.

Trzeba więc zwrócić uwagę, że doskonale odwzorowywanie walorów ludzkiego głosu przez technologie AI uruchamia automatyzmy osobowego postrzegania istoty mówiącej, nawet jeśli jest ona robotem. Pojęcia „ja” i „inny” transponujemy bowiem na walory głosu, ponieważ różne głosy reprezentują różne osoby (Sutton i in., 2019). Skoro posiadamy świadomość jako byty mówiące, „obdarzamy” nią również asystenta budującego z nami relację za pośrednictwem głosu. Tym bardziej, że cechy głosu AI nie są przypadkowe, ale wynikają ze specyfiki relacji międzyludzkich i to właśnie one stanowią bazę projektowania technologii sztucznej inteligencji.

Słuchacz nie może stłumić swoich naturalnych reakcji na mowę, niezależnie od źródła. Ludzie wyciągają wnioski na temat głosów opartych na technologii i podejmują odpowiednie zachowanie, stosując te same

---

2 Eksperyment dotyczył asystenta Microsoft Cortana (Cambre i Chinmay, 2019).

zasady i skróty, których używają podczas interakcji z ludźmi. Technologie te, podobnie jak mowa innych ludzi, aktywują wszystkie części mózgu związane z interakcjami społecznymi (Nass i Brave, 2005, s. 4 w: Edwards i in., 2019, s. 358).

Co ważne, relacje w aspekcie naturalnym, praktycznym czy społecznym wywołują różne emocje związane na przykład z dobrym samopoczuciem fizycznym, osiągnięciami performatywnymi i poczuciem własnej wartości (Archer, 2021). To istotny aspekt komunikacji człowieka z asystentami wirtualnymi, ponieważ roboty są uczone analizowania emocji ludzi wyrażanych w dialogu za pomocą sygnałów werbalnych i niewerbalnych. Nie tylko zatem dekodują stan emocjonalny podmiotu mówiącego, ale także symulują swoje adekwatne reakcje emocjonalne (Devillers, 2021).

### Świadomość głosowej odrębności

W technologiach AI następuje personalizacja głosu, czyli modulowanie jego parametrów, by imitował czyjś konkretny głos także w aspekcie wyrażania emocji. Zatem asystenci stają się coraz bardziej realistycznymi (*life-like*) partnerami komunikacyjnymi (Guzman i Lewis, 2019, s. 73). To istotne z uwagi na fakt, że emocjonalność głosu AI (Seaborn i in., 2021) wzmacnia potencjał wzajemności w komunikacji. Tym samym użytkownik ma poczucie, że asystent świadomie reguluje emocje i dostosowuje je do nastroju interlokutora. Pojawia się tu problematyczna kwestia głosowej odrębności człowieka, którego sklonowanym głosem posługuje się asystent.

Najnowsza technologia WALL-E pozwala kopiować własny głos i prowadzić dialog ze swoim głosowym awatarem (cyfrową kopią własnego głosu). Ten fakt nie może pozostawać bez wpływu na poczucie tożsamości człowieka, skoro nasz prawdziwy głos to, jak pisze Adam Lipszyc (2022), „samo sedno tego, kim jesteśmy” (s. 47). Człowiek jest dawcą – naturalnym właścicielem głosu. Personifikowany głos AI w dialogu z właścicielem tego samego głosu zmienia porządek odczytywania „ja” (dawcy) w relacji do „innego”. Nasze „ja” rozwija się bowiem z poczucia tożsamości (Archer, 2021). Jako ludzie posiadamy świadomość swojej osobowej odrębności. Każdego dnia wykorzystujemy język do stawiania pytań wewnętrznych, adresowanych do samego siebie, a także do innych. W pierwszym przypadku jesteśmy zarówno mówiącym, jak i słuchającym, prowadzimy coś na kształt wewnętrznego dialogu. Jesteśmy też w stanie klasyfikować głosy, które słyszymy. Mladen Dolar (2006) konstatował, że słyszenie siebie mówiącego jest potrzebne do wytworzenia minimalnej

formy jaźni, zaś głos jest „odnoszącym się do samego siebie”, „auto-refleksyjnym” odruchem, ujawniającym się jako auto-afekt (Location 632). Nasz własny głos jest nam bliski, stanowi „część nas”. Dialog ze spersonalizowanym głosem AI, dla którego sami jesteśmy „dawcą”, odsyła nas do narcystycznego wsłuchiwania się w głos, który do nas powraca. Mityczny Narcyz słyszał jednak głos nimfy Echo powtarzającej jedynie jego własne słowa (Dolar, 2006). W tej specyficznej sytuacji komunikacyjnej *quasi*-dialogu Narcyz rozmawiał z samym sobą, ale na poziomie semantyki fraz nie następowało otwieranie kolejnych tur dialogu. Dopóki Narcyz nie zmienił formuły, dopóty Echo powtarzała wciąż jego słowa. Syntetyczny głos asystenta umożliwia natomiast dialog „dawcy” z samym sobą, nie zaś duplikowanie wcześniej wybrzmiałych fraz. Staje się rozmową, w której głos dawcy (generowany przez asystenta) sygnalizuje jego tożsamość (poprzez jego cechy brzmieniowe), ale jednocześnie funkcjonuje poza jego świadomymi wyborami werbalnymi. Asystent jako oddzielny „byt” korzysta ze swoich ścieżek dialogu, ale jednocześnie mówi głosem „dawcy”. Ta sytuacja komunikacyjna powoduje dysonans poznawczy: człowiek słyszy swój głos w wydaniu asystenta, ale nie ma wpływu na jego emisję.

Oznacza to technologiczny „zamach” na uświadamianie sobie przez człowieka swojego „ja”, ponieważ, jak twierdzi Thomas Ogden, „głos jest doświadczeniem ‘ja’ powstającym w akcie mówienia” (Ogden, 1998, s. 445, za: Lipszyc, 2022, s. 39). Głos ujawnia się w relacji między podmiotem a „Innym” stanowiąc, według słów Dolara, jednocześnie nieuchwytnie „jądro naszej podmiotowości”, jak i „centrum naszych społecznych więzi” (Dolar, 2006, s. 14, za: Lipszyc, 2022, s. 49). Ponieważ każdy komunikujący identyfikuje swoje własne atrybuty dzięki interakcjom (Goffman, 1959), klonowanie głosu burzy porządek dialogu z „Innym”, umożliwiając dialog z samym sobą w znaczeniu interakcji.

Ten aspekt niemożności identyfikowania jedynie siebie samego poprzez atrybuty własnego głosu pogłębia fakt, że tożsamość jest powiązana z osobowością – złożoną właściwością umysłu, jaką jest perspektywa pierwszej osoby. Pozwala nam ona postrzegać nasze ciało i stany mentalne jako własne (Baker, 2000), ponieważ jest dopełniana refleksyjnością niezbędną w definiowaniu osobistej i społecznej tożsamości, zjawisk wewnętrznych oraz zewnętrznych – czyli świadomością. Biologiczne uwarunkowania tego fundamentalnego stanu psychicznego na przestrzeni dziejów były gatunkowo zróżnicowane. Możemy zatem zakładać, że pewna podstawa świadomości jest możliwa do replikowania. Czy jednak oznacza to, że świadomość jako taka może być zaimplementowana w materii nieożywionej? Filozof Markus Gabriel (2021) twierdzi, że zbliżenie się robota do stanu świadomości byłoby czystym zbiegiem okoliczności, zaś „przekonanie, że każdy istniejący robot lub komputer jest świadomy,

jest równie racjonalne jak to, że Droga Mleczna lub burza piaskowa na pustyni Atakama są świadome” (s. 61). Ten aspekt znajduje odzwierciedlenie m.in. w modelu dialogu, który jest w stanie prowadzić asystent AI. Zdaniem Rafała Garbacza, specjalisty w zakresie technologii głosowych, Rubikon zostanie przekroczony wtedy, gdy asystent zyska poziom świadomości umożliwiający mu samodzielne inicjowanie tematów w dialogu i będzie to robił naturalnie, jak człowiek. Obecnie wciąż brakuje mu kreatywności i spontaniczności, by podejmować różne wątki i dostosowywać odpowiedzi intuicyjnie i zgodnie z naszymi intencjami (RG, 2023).

W tym kontekście nowe światło na kształtowanie parametrów głosowych asystentów rzucają badania wskazujące, że takie ograniczenia głosu AI jak brak ciągłości tonu czy natężenia nie zawsze są postrzegane przez użytkowników w kategorii mankamentu (Cabral, Cowan, Zibrek i McDonnell, 2017). Niektórzy badacze (Cambre i Chinmay, 2019 za: Chan, 2021) sugerują nawet rozważenie celowego kreowania głosów w wirtualnych interfejsach, które są wyraźnie „nie-ludzkie” (*non-human*), aby cechy głosu asystenta wspierały „osobowość” robota, nie zaś ją podważały (Aylett, Sutton i Vazquez-Alvarez, 2019 za: Chan, 2021). Wydaje się zatem, że wraz z nurtem klonowania głosu naturalnego przez mechanizmy AI rozwija się także nurt klarownego i celowego wypuklania przymiotów głosu robotycznego. Bezspornie ma to związek z intuicyjną potrzebą ochrony cech głosu naturalnego jako matecznika tożsamości i wyjątkowości człowieka.

## Podsumowanie

Tożsamość jednostek i zbiorowości kształtuje się w relacjach. Jak pisze Pierpaolo Donati (2021)

hybrydyzacja tożsamości ludzi i organizacji społecznych (...) polega na tym, że podmioty te zmieniają swoją relacyjność w takim stopniu, w jakim kulturowe procesy „przekraczają” osobę ludzką i modyfikują jej sposoby odnoszenia się do siebie, do innych i do świata (s. 217).

Diagnoza problemu głosu naturalnego jako elementu tożsamości w hybrydycznej rzeczywistości relacji człowiek – asystent głosowy obejmująca aspekty wzrastającej roli komunikacyjnej asystentów głosowych AI, transponowania przez użytkowników tożsamości głosowej na asystenta oraz świadomości głosowej odrębności – pozwala wnioskować, że tożsamość człowieka migruje w przestrzeń hybrydycznych relacji z asystentami z powodu nowej logiki tych relacji. Relacje człowiek – asystent z definicji



nie mogą bowiem pozostać zamknięte w samoreferencyjności właściwej sztucznej inteligencji. Wchodząc w wielokrotne interakcje z asystentami, ludzie jako istoty społeczne intuicyjnie włączają pewne aspekty naturalnej komunikacji do swojego kontaktu z technologią. Pogłębia tę tendencję fakt, że głos asystenta dzięki udoskonalanym algorytmom staje się nośnikiem emocji kompatybilnych z emocjami użytkownika. Przejęcie przez technologię AI dystynktywnych cech głosu jako audialnych linii papilarnych neguje kulturową spójność głosu z ciałem emitującego go człowieka, warunkującą jego tożsamość. Możliwość klonowania głosu i prowadzenia dialogu z asystentem mówiącym głosem interlokutora burzy antropologiczny porządek odrębności własnego oraz cudzego głosu.

#### BIBLIOGRAFIA

- Aeschlimann, S., Bleiker, M., Wechner, M. i Gampe, A. (2020). Communicative and social consequences of interactions with voice assistants. *Computers in Human Behavior*, 112, Article 106466. DOI: 10.1016/j.chb.2020.106466
- Archer, M.S. (2021). Friendship Between Human Beings and AI Robots? W: J. von Braun i in. (red.), *Robotics, AI, and Humanity*. Cham: Springer, 177–191. DOI: 10.1007/978-3-030-54173-6\_16177-188.
- Aylett, M.P., Sutton, S.J. i Vazquez-Alvarez, Y. (2019). The Right Kind of Unnatural: Designing a Robot Voice. W: *Proceedings of the 1st International Conference on Conversational User Interfaces* (Dublin, Ireland) (CUI '19). Association for Computing Machinery, New York, NY, USA, Article 25, 2 pages. DOI: 10.1145/3342775.3342806.
- Baker, L.R. (2000). *Persons and bodies: A constitution view*. Cambridge: Cambridge UP.
- Cabral, J.P., Cowan, B.R., Zibrek, K. i McDonnell, R. (2017). The Influence of Synthetic Voice on the Evaluation of a Virtual Character. *INTERSPEECH*, 229–233. DOI: 10.21437/Interspeech.2017-325.
- Cambre, J. i Chinmay, K. (2019). *One Voice Fits All?: Social Implications and Research Challenges of Designing Voices for Smart Devices*. Proceedings of the ACM on Human-Computer Interaction Volume 3 Issue CSCW November 2019, 1–19. DOI: 10.1145/3359325.
- Chan, S.W.T., Gunasekaran, T.S., Pai, Y.S., Zhang, H. i Nanayakkara, S. (2021). Kin Voices: Using Voices of Friends and Family in Voice Interfaces. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW2, Article 446 (October 2021), 25 pages. DOI: 10.1145/3479590.
- Chin, C. i Robison, M. (2020). *How AI bots and voice assistants reinforce gender bias*, November 23, 2020. Pozyskano z: <https://www.brookings.edu/research/how-ai-bots-and-voice-assistants-reinforce-gender-bias> (dostęp: 28.01.2023).

- Coeckelbergh, M. (2011). Humans, animals, and robots: A phenomenological approach to human-robot relations. *International Journal of Social Robotics*, nr 3(2), 197–204. DOI: 10.1007/s12369-010-0075-6).
- Devillers, L. (2021). Human–Robot Interactions and Affective Computing: The Ethical Implications. W: J. von Braun i in. (red.), *Robotics, AI, and Humanity*. Cham: Springer, 205–211. DOI: 10.1007/978-3-030-54173-6\_16177-188.
- Dolar, M. (2006). *Voice and Nothing More*. MIT Press. Kindle Edition.
- Donati, P. (2021). Impact of AI/Robotics on Human Relations: Co-evolution Through Hybridisation. W: J. von Braun i in. (red.), *Robotics, AI, and Humanity*. Cham: Springer, 213–227. DOI: 10.1007/978-3-030-54173-6\_16177-188.
- Edwards, C., Edwards, A., Stoll, B., Lin, X. i Massey, N. (2019). Evaluations of an artificial intelligence instructor’s voice: Social Identity Theory in human-robot interactions. *Computers in Human Behavior*, 90, 357–362.
- Gabriel, M. (2021). Could a Robot Be Conscious? Some Lessons from Philosophy. W: P. Donati (red.), *Impact of AI/robotics on human relations: co-evolution through hybridisation*. *Robotics, AI, and Humanity*. Cham: Springer, 213–227.
- Goffman, E. (1959). *The Presentation of Self in Everyday Life*. New York: Anchor Books.
- Guzman, A.L. i Lewis, S.C. (2019). Artificial intelligence and communication: A human-machine communication research agenda. *New Media Society*, nr 22 (1), 70–86. DOI: 10.1177/1461444819858691.
- Juniper Research. (2018). Pozyskano z: <https://www.juniperresearch.com/press/digital-voice-assistants-in-use-to-8-million-2023> (dostęp: 17.02.2023).
- Kreiman, J., Vanlancker-Sidtis, D. i Gerratt, B.R. (2003). *Defining and measuring voice quality*. W: *Proceedings of the Conference on Voice Quality: Functions, Analysis, and Synthesis*. ISCA, 115–120. Pozyskano z: [https://www.isca-speech.org/archive\\_open/voqual03/voq3\\_115.html](https://www.isca-speech.org/archive_open/voqual03/voq3_115.html) (dostęp: 17.12.2019).
- Lipszyc, A. (2022). The Shrimp-Mirror-Stitch, or Voice in Psychoanalysis, *Eidos. A Journal for Philosophy of Culture*, nr 2(6), 37–50, DOI: 10.14394/eidos.jpc.2022.0014.
- McCarthy, J. (2007). *What is Artificial Intelligence?* Pozyskano z: <http://www-formal.stanford.edu/jmc/> (dostęp: 25.01.2023).
- McTear, M.F., Callejas, Z. i Griol, D. (2016). *The Conversational Interface: Talking to Smart Devices*. Springer, Switzerland.
- Nass, C. (2005). *Machine voices*. Pozyskano z: <http://www.pbs.org/speak-ahead/technology/voiceinterface/#> (dostęp: 09.02.2023).
- Nass, C. i Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship*. Boston Review.

- Nass, C. i Lee K.M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology Applied*, nr 7(3), 171–181.
- Nass, C. i Moon, Y. (2000). Machines and mindlessness: social responses to computers. *Journal of Social Issues*, nr 56(1), 81–103. DOI:10.1111/0022-4537.00153.
- Nass, C., Steuer, J. i Tauber, E.R. (1994). Computers Are Social Actors. W: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, Massachusetts, USA) (CHI '94). Association for Computing Machinery, New York, NY, USA, 72–78. DOI: 10.1145/191666.191703.
- Ogden, T. (1998). A Question of Voice in Poetry and Psychoanalysis. *The Psychoanalytic Quarterly*, nr 67, 426–448. DOI: 10.1080/00332828.1998.12006050.
- Ostrowski, A., Fu, J., Zygoras, V., Park, H. i Breazeal, C. (2022). Speed Dating with Voice User Interfaces: Understanding How Families Interact and Perceive Voice User Interfaces in a Group Setting. *Frontiers in Robotics and AI*. 10.3389/frobt.2021.73092.8.
- Pitardi, V. i Marriott, H.R. (2021). Alexa, she's not human but... Unveiling the drivers of consumers' trust in voice based artificial intelligence. *Psychology & Marketing*, nr 38(4), 626–642.
- Reeves, B. i Nass, C.I. (1996). *The Media Equation*. Stanford, CA: CSLI Publications.
- Schröder, W.M. (2021). Robots and Rights: Reviewing Recent Positions, in Legal Philosophy and Ethics. W: J. von Braun i in. (red.), *Robotics, AI, and Humanity*. Cham: Springer, 191–202. DOI: 10.1007/978-3-030-54173-6\_16177-188.
- Seaborn, K., Miyake, N.P., Pennefather, P. i Otake-Matsuura, M. (2021). Voice in human-agent interaction: a survey. *ACM Computing Surveys* (CSUR), 54(4), 1–43.
- Smith, S.V. (2015). An NPR reporter raced a machine to write a news story. Who won? NPR, 20 May. Pozyskano z: <https://www.npr.org/sections/money/2015/05/20/406484294/an-nprreporter-raced-a-machine-to-write-a-news-story-who-won> (dostęp: 30.01.2023).
- Sutton, S.J., Foulkes, P., Kirk, D. i Lawson, S. (2019). Voice as a Design Material: Sociophonetic Inspired Design Strategies in Human-Computer Interaction. W: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) (CHI '19). Association for Computing Machinery, New York, NY, USA, 1–14. DOI: 10.1145/3290605.3300833.
- Tarnóczy, T.H. (1949). The Speaking Machine of Wolfgang von Kempelen. *The Journal of the Acoustical Society of America*, nr 21(461), <https://asa.scitation.org/doi/abs/10.1121/1.1917078>, (dostęp: 27.01.2023).

- Wagner, K. i Schramm-Klein, H. (2019). *Alexa, Are You Human? Investigating Anthropomorphism of Digital Voice Assistants – A Qualitative Approach*. ICIS Proceedings. 7. Pozyskano z: [https://aisel.aisnet.org/icis2019/human\\_computer\\_interact/human\\_computer\\_interact/7](https://aisel.aisnet.org/icis2019/human_computer_interact/human_computer_interact/7), (dostęp: 28.01.2023).
- Wang, Ch., Chen, S., Wu, Y., Zhang, Z., Zhou, Shujie, L., Chen, Z., Liu, Y., Wang, H., Li, J., He, L., Zhao, S. i Wei, F. (2023). *VALL-E Neural Codec Language Models are Zero-Shot Text to Speech Synthesizers*. arXiv:2301.02111 [cs.CL].
- Zellou, G., Cohn, M. i Ferenc S.B. (2021). Age-and gender-related differences in speech alignment toward humans and voice-AI. *Frontiers in Communication*, vol. 5. DOI: 10.3389/fcomm.2020.600361.

**Grażyna Stachyra** – dr hab., profesor zwyczajny UMCS w Katedrze Komunikacji Medialnej Instytutu Nauk o Komunikacji Społecznej i Mediach. Bada nowe formy komunikacji audialnej i emisję głosu w mediach audiowizualnych. Autorka m.in. *The Radio Plays Games* („European Journal of Communication”, Vol. 27(3)2012); *Radio in the workplace. A liminal medium between work and leisure* („Media Culture & Society”, Vol. 37(2)2015).