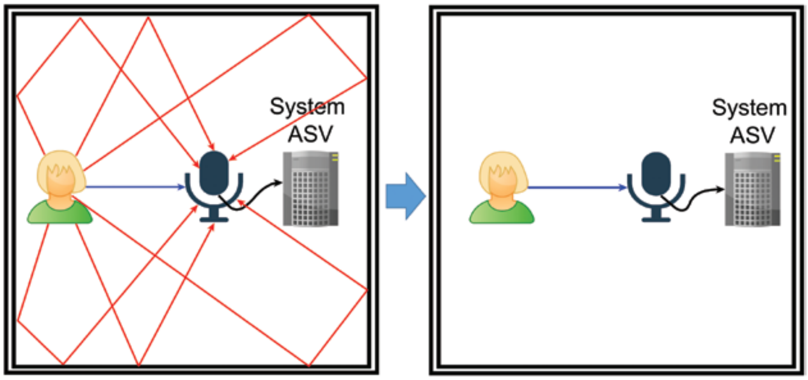
Dr inż. Marcin Witkowski

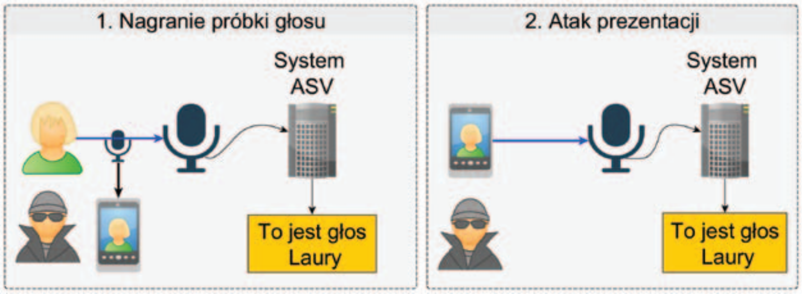
Rozmowa z autorem pracy:
„Robust speaker verification with reverberation suppression and spoofing detection”

Co sprawiło, że zainteresował się Pan akurat tymi aspektami bezpieczeństwa w telekomunikacji?

W początkowych latach w Zespole Przetwarzania Sygnałów AGH w Katedrze Elektroniki miałem przyjemność pracować w projekcie poświęconym tworzeniu rozwiązania umożliwiającego weryfikację użytkowników na podstawie głosu. W tym czasie poznałem podstawowe problemy naukowe związane z wdrażaniem tej technologii, co nakreśliło moją ścieżkę badań. W ciągu ostatnich 15 lat można było zaobserwować dwa przełomy w metodach rozpoznawania mówców wyłącznie na podstawie sygnału audio, co pozwoliło osiągnąć satysfakcjonującą skuteczność w warunkach laboratoryjnych. Niestety, gdy warunki akustyczne podczas rejestracji użytkownika są inne niż te w trakcie weryfikacji, skuteczność systemu znacznie spada. Innym aspektem, branym pod uwagę podczas wdrażania, jest bezpieczeństwo systemu rozumiane przez zdolność do automatycznego wykrywania ataków. W mojej pracy zaadresowałem obydwa problemy proponując metodę detekcji ataków prezentacji, czyli takich, w których głos użytkownika jest nagrany, a następnie odtworzony przez atakującego w celu uzyskania dostępu. Zaproponowałem również metodę usuwającą pogłos z już zarejestrowanego sygnału audio, mającą na celu zniwelowanie różnic w warunkach akustycznych pomiędzy rejestracją a testem. Celem mojej pracy doktorskiej było więc opracowanie rozwiązań umożliwiających szersze zastosowanie technologii biometrii głosowej.



Redukcja pogłosu



Schemat ataku prezentacji

Czy ostatnie sukcesy na polu ogólnodostępnych platform AI mają przełożenie na możliwości ataku jak i obrony przed nim?

Oczywiście trwają wzmożone prace zarówno nad coraz bardziej wiernym odtworzeniem mowy i używaniem jej do prób oszukania systemów weryfikacji mówców jak i nad wykrywaniem tego typu ataków. Skuteczność wykrywania ataków można śledzić obserwując publikacje z konferencji Interspeech, a w szczególności wyniki konkursów ASVspoof Challenge czy Voice Privacy Challenge. Trzeba mieć również na uwadze, że istnieją już technologie oparte na głębokich sieciach neuronowych, umożliwiające wygenerowanie (syntezę) wypowiedzi dowolnej osoby, dla której posiadamy nagrany krótki fragment audio. Nawiasem mówiąc, znam absolwentów AGH, którzy zajmują się tą tematyką i osiągają niesamowite efekty w swych pracach. Nie sądzę jednak, by ostatnie sukcesy platform AI takich jak ChatGPT miały dziś bezpośredni wpływ na tę dziedzinę. Na ten moment ogólnodostępne modele językowe, określane często jako sztuczna inteligencja, mogą kierować do repozytoriów z kodem źródłowym dla relatywnie starych modeli (sprzed 2 lat) oraz generować tekst w stylistyce konkretnej osoby. Można się jednak spodziewać, że popyt na komunikację głosową z modelami językowymi przyspieszy pośrednio rozwój zarówno metod konwersji głosu jak również ochrony przed nimi.

W jaki sposób zasoby Cyfronetu przysłużyły się realizacji części Pana badań związanej z obliczeniami i ich analizą?

W ramach moich prac korzystałem zarówno z zasobów umożliwiających zrównoleglenie obliczeń z użyciem kart graficznych jak i bez ich wykorzystania. W szczególności w eksperymentach poświęconych opracowaniu metody usuwania pogłosu korzystałem z superkomputera Prometheus, zarówno w celu usunięcia pogłosu jak i oceny obiektywnej jakości usuwania pogłosu. Zrównoleglenie polegało na przetwarzaniu wielu plików audio jednocześnie. Karty graficzne dostępne w ramach superkomputera były wykorzystywane do treningu modeli do ekstrakcji cech mówców (ang. speaker embedding) oraz w dalszej kolejności do ekstrakcji cech w celu ewaluacji weryfikacji mówcy. Brak dostępu do infrastruktury obliczeniowej PLGrid skutkowałby zwiększeniem czasu obliczeń w zależności od eksperymentu od kilkunastu do kilkuset razy, co uniemożliwiłoby mi zarówno publikację wyników prac badawczych na znaczących konferencjach jak i ukończenie badań na poczet pracy.

Co mógłby Pan doradzić osobom, które dopiero zaczynają studia doktoranckie? Na co te osoby powinny zwrócić największą uwagę?

Badania naukowe są z natury trudne, gdyż wymagają poruszania się w nieznanych tematach, gdzie odpowiedzi należy szukać samodzielnie ale nie samotnie. Bardzo duży wpływ na sukces bądź porażkę ma promotor oraz dostęp do warsztatu badawczego. Dobra relacja z promotorem, który jest w stanie wspomóc radą czy konstruktywną rozmową w obszarze badawczym, jest w moim rozumieniu kluczowa. Dodatkowo należy zwrócić uwagę na to, by podejmując temat mieć dostęp do niezbędnej infrastruktury obliczeniowej. Przykładowo, chcąc wykonywać prace oparte na sieciach neuronowych, niezbędne jest posiadanie dostępu do danych jak i maszyn umożliwiających przeprowadzenie eksperymentów na dużą skalę, takich jak te udostępniane przez Cyfronet.