

Piotr KŁOSOWSKI, Adam DUSTOR, Jacek IZYDORCZYK
Politechnika Śląska, Instytut Elektroniki

ZASTOSOWANIE ROZPOZNAWANIA MÓWCY W AUTOMATYCZNEJ TRANSLACJI MOWY TYPU SPEECH-TO-SPEECH¹

Streszczenie. Przedstawiony artykuł dotyczy zagadnień związanych z funkcjonowaniem systemów automatycznej translacji mowy ciągłej. W systemach tych wykorzystuje się techniki przetwarzania języka naturalnego realizowane z wykorzystaniem algorytmów automatycznego rozpoznawania mowy, automatycznej translacji tekstów oraz zamiany tekstu na mowę za pomocą syntezy mowy. W artykule zaproponowano także metodę usprawnienia procesu automatycznej translacji mowy przez zastosowanie algorytmów automatycznej identyfikacji mówcy, pozwalających na automatyczną segmentację mowy pochodzącej od różnych mówców.

Słowa kluczowe: rozpoznawanie mówcy, rozpoznawanie mowy, translacja mowy, translacja tekstu, synteza mowy

SPEAKER RECOGNITION APPLICATION IN AUTOMATIC SPEECH-TO-SPEECH TRANSLATION

Summary. This paper concerns the machine translation of continuous speech. These systems use machine language processing techniques implemented using algorithms of automatic speech recognition, automatic text translation and text-to-speech conversion using speech synthesis.

Keywords: speaker recognition, speech recognition, speech translation, text translation, speech synthesis

¹ Artykuł powstał dzięki dofinansowaniu przyznanemu przez Narodowe Centrum Badań i Rozwoju w ramach grantu nr POIG.01.03.01-24-107/12 (Opracowanie innowacyjnej metody identyfikacji mówcy dla podniesienia stopnia bezpieczeństwa systemów teleinformatycznych).

1. Wstęp

Zakład Telekomunikacji Instytutu Elektroniki na Wydziale Automatyki, Elektroniki i Informatyki Politechniki Śląskiej od wielu lat prowadzi badania w dziedzinie cyfrowego przetwarzania sygnałów i jego zastosowania w telekomunikacji. Jeden z najważniejszych obszarów badań stanowi cyfrowe przetwarzanie sygnału mowy, obejmujące zagadnienia analizy i syntezy sygnału mowy oraz rozpoznawania mowy. W ostatnich latach obszar badawczy został poszerzony o zagadnienia automatycznej translacji mowy.

Dziedzina automatycznej translacji mowy typu speech-to-speech (ang. SSMT – *Speech-to-Speech Machine Translation*) wpisuje się w ugruntowany już obszar badań nad przetwarzaniem sygnału mowy oraz języka naturalnego. Jest to dziedzina o ogromnym znaczeniu, z którą wiąże się ogromne nadzieje, ponieważ dotyczy podstawowych problemów i potrzeb współczesnego społeczeństwa informacyjnego, takich jak komunikacja między ludźmi oraz dostęp do informacji w różnych językach, co ma kluczowe znaczenie w dzisiejszym coraz bardziej zglobalizowanym świecie [1].

2. Automatyczna translacja mowy i jej zastosowania

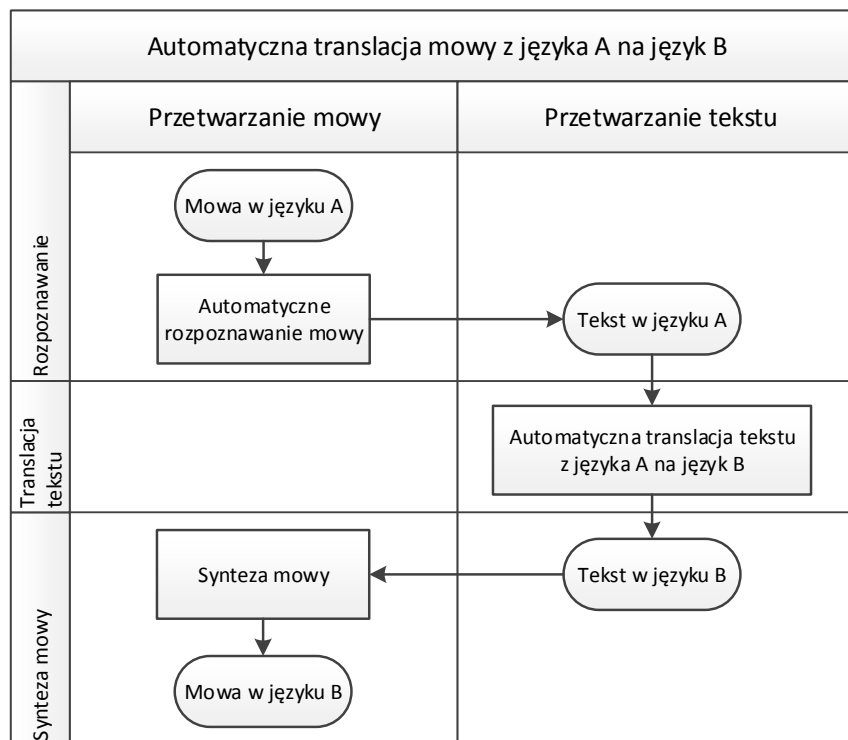
Współcześnie rozróżnia się ponad 7000 języków [2]. Bogactwo i różnorodność języków we współczesnym świecie uznaje się za dorobek i ukoronowanie ewolucji człowieka [3]. Język dla danej grupy etnicznej ma ogromne znaczenie i stanowi integralny element danej społeczności. Jest dziedzictwem kulturowym oraz decyduje o odrębności danej grupy. Dodatkowo obserwuje się tendencję do zanikania najmniej popularnych języków na rzecz tych używanych powszechnie.

Jeśli obecna technika byłaby w stanie dostarczyć technologii automatycznego tłumaczenia mowy na inne języki, pojawiłaby się możliwość łatwego dostępu do informacji w dowolnym języku oraz możliwość komunikacji między użytkownikami w wielu językach. Obecnie translacja tekstu na inne języki dokonywana jest zwykle przez człowieka. Ogromne zapotrzebowanie na tego typu usługi oraz różnorodność języków sprawia, że liczba tłumaczy jest wciąż niewystarczająca, a proces ten jest bardzo kosztowny.

Dobrym przykładem może być Unia Europejska, która uznała znaczenie różnorodności językowej, jako niezbywalnego elementu dziedzictwa kulturowego narodów. Doprowadziło to do przyjęcia zasady równości wszystkich 23 języków, używanych w 27 krajach, członków Unii Europejskiej [4]. Założenie, że wszystkie 23 języki są równie ważne, w konsekwencji powoduje, że cała komunikacja w administracji Unii Europejskiej, np. wypowiedzi deputowanych w Parlamencie Europejskim lub dokumenty, którymi posługują się unijni urzędnicy

oraz instytucje Unii Europejskiej muszą być tłumaczone na 23 języki [5]. Jest to bardzo kosztowne przedsięwzięcie, które pochłania corocznie ok. 1 mld euro tylko na wykonywanie niezbędnych tłumaczeń w tradycyjny sposób.

Zastosowanie do wykonywania tłumaczeń nowoczesnej technologii w postaci automatycznej translacji mowy typu speech-to-speech może odegrać w przyszłości kluczową rolę w komunikacji głosowej oraz dostępie do treści mówionych, zapewniając odpowiednią jakość tłumaczenia w przystępnej cenie.



Rys. 1. Schemat blokowy systemu automatycznej translacji mowy
Fig. 1. Block diagram of speech-to-speech translation

Systemy automatycznej translacji mowy wykorzystują techniki przetwarzania języka naturalnego realizowane z wykorzystaniem algorytmów automatycznego rozpoznawania mowy, rozpoznawania mówcy, automatycznej translacji tekstów oraz zamiany tekstu na mowę za pomocą syntezy mowy [6]. Typowy system automatycznej translacji mowy może się składać z następujących modułów funkcjonalnych:

- modułu automatycznego rozpoznawania mowy ciągłej, pozwalającego zamienić mowę na tekst;
- modułu automatycznej translacji tekstu z języka źródłowego do języka docelowego;
- modułu syntezy mowy, pozwalającego zamienić przetłumaczony wynikowy tekst na sygnał mowy w języku docelowym.

Schemat blokowy typowego systemu automatycznej translacji mowy został przedstawiony na rysunku 1. W dalszej części artykułu zostaną przedstawione szczegóły funkcjonowania

poszczególnych modułów oraz całego systemu automatycznej translacji mowy. Zaprezentowano także ideę usprawnienia procesu automatycznej translacji mowy poprzez zastosowanie algorytmów automatycznej identyfikacji mowy, pozwalających na automatyczną segmentację mowy pochodzącej od różnych mówców.

3. Wykorzystanie automatycznego rozpoznawania mowy w automatycznej translacji tekstów

Systemy rozpoznawania mowy, z punktu widzenia zakresu zastosowań oraz charakteru danych wejściowych, można podzielić na kilka kategorii [7]. Głównym wyróżnikiem podziału i złożoności systemu rozpoznawania mowy jest to, czy sygnał mowy reprezentuje pojedyncze izolowane słowa lub czy jest sygnałem mowy ciągłej. W mowie ciągłej, a z tym przypadkiem mamy do czynienia w systemach automatycznej translacji mowy, jednym z podstawowych zagadnień jest trudność w określeniu początku i końca poszczególnych słów lub innej jednostki znaczeniowej, np. frazy lub zdania [8].

W systemach ograniczonego rozpoznawania mowy zakłada się, że sygnał mowy nie zawiera wszystkich niezbędnych informacji wystarczających do bezpośredniego rozpoznania komunikatu. Z tego względu stosuje się w pewnym stopniu dodatkowe informacje językowe, które wraz z ograniczonym słownikiem oraz warunkami odnośnie do sposobu wymowy powinny zapewnić wystarczającą skuteczność rozpoznawania. Systemy ograniczonego rozpoznawania i rozumienia mowy korzystają z wszelkich możliwych dodatkowych informacji językowych, umożliwiających rozpoznanie i zrozumienie treści mowy. W systemach tego typu nie jest ważne rozpoznanie każdego fonemu lub słowa. O skuteczności systemu decyduje końcowy efekt rozpoznania treści wypowiedzi.

Systemy nieograniczonego rozpoznawania mowy ciągłej zakładają brak dodatkowych informacji lingwistycznych i z tego względu ich realizacja wydaje się być najbardziej skomplikowana. Uzupełnienie procedur rozpoznawania, w systemach nieograniczonego rozpoznawania mowy, o pełne informacje lingwistyczne pozwoli na realizację systemów pełnego rozpoznawania i rozumienia mowy.

Istnieje kilka metod projektowania systemów automatycznego rozpoznawania mowy. Najważniejsze z nich to:

- metody akustyczne, polegające głównie na wykorzystaniu parametrów fonetyczno-akustycznych, służących do konstruowania obrazów i wzorców, najczęściej w systemach probabilistycznych,
- metody oparte na znajomości wytwarzania mowy przez człowieka, polegające na analizie głównych procesów artykulacji poszczególnych jednostek językowych,

- metody oparte na bionicznej koncepcji rozpoznawania, czyli na modelowaniu procesu rozpoznawania mowy przez człowieka,
- metody uwypuklające proces rozumienia mowy na podstawie pewnych charakterystycznych cech prozodycznych ważnych dla rozpoznania, jak: np. transjenty (przejścia międzyfonemowe formantów), dźwięczność (bezdźwięczność) spółgłosek, charakter zmian intonacji itp.

W rzeczywistych systemach automatycznego rozpoznawania mowy najczęściej mamy do czynienia z pewną kombinacją wymienionych czterech metod, z tym że jedna z nich pełni zwykle funkcję dominującą. W większości skutecznie działających systemach automatycznego rozpoznawania mowy najczęściej mają zastosowanie dwie pierwsze metody.

Podstawowym etapem procesu rozpoznawania jest analiza na poziomie akustycznym, w wyniku której następuje ekstrakcja parametrów akustycznych. Drugim poziomem rozpoznawania jest poziom ekstrakcji parametrów fonetycznych, niezbędnych do rozpoznawania fonemów. Na podstawie analizy akustycznej następuje równoległe wydzielenie charakterystyk prozodycznych, jak intonacja, akcent, rytm, które pozwalają na określenie głównych struktur lingwistycznych. Kolejnym poziomem jest segmentacja i klasyfikacja fonetyczna, pozwalająca na ustalenie określonych ciągów fonetycznych. Ciągi fonetyczne pozwalają na realizację procedury porównania słów ze słowami ze słownika. Hipotetyczne słowa (lub inne jednostki językowe) kończą proces rozpoznawania.

4. Automatyczna translacja tekstu

Tłumaczenie maszynowe polega na automatycznym tłumaczeniu tekstu lub mowy z jednego języka na drugi i jest jednym z najważniejszych zastosowań przetwarzania języka naturalnego (ang. *Natural Language Processing*), dziedziny z pogranicza informatyki i lingwistyki [9]. Zagadnienie tłumaczenia maszynowego nie jest pomysłem nowym. Pierwsze odnotowane rozwiązania tego typu pochodzą z XVII wieku, jednak dopiero rozwój techniki cyfrowej i zastosowanie komputerów pozwoliło na znaczący rozwój tej dziedziny. Przez ostatnich kilkadziesiąt lat w prace nad tłumaczeniem maszynowym zainwestowano olbrzymie fundusze i pomimo wieloletnich wysiłków licznych ośrodków badawczych problem nie został rozwiązany w stopniu zadowalającym, pozwalającym na skuteczne dokonywanie dowolnych tłumaczeń. Można wyróżnić następujące rodzaje systemów tłumaczenia maszynowego:

- tłumaczenie wspomagane maszynowo (ang. *Machine Aided Human Translation*),
- tłumaczenie wspomagane przez człowieka (ang. *Human Aided Machine Translation*),
- tłumaczenie całkowicie maszynowe (ang. *Fully Automated Machine Translation*).

W systemach automatycznej translacji mowy znajduje zastosowanie najczęściej całkowicie maszynowe tłumaczenie tekstu. Polega ono na tym, że tłumaczeniu podlega tekst źródłowy i bez jakiegokolwiek ingerencji człowieka generowany jest tekst w języku wynikowym. Całkowicie maszynowe tłumaczenie tekstów daje obecnie najgorsze wyniki, jeśli chodzi o dokładność tłumaczeń. Najczęściej tekst wynikowy jest dość niskiej jakości, jednak wystarczającej do pewnych zastosowań, jak na przykład przeglądanie stron WWW lub czytanie poczty elektronicznej. W niektórych zastosowaniach najważniejszy jest szybki dostęp do informacji, a nie dokładność tłumaczenia. Jeżeli zadowalający będzie przybliżony przekład tekstu lub dokonywane są tłumaczenia jedynie w wąskiej dziedzinie, to technologia tłumaczenia maszynowego osiągnęła już wystarczający poziom, by znaleźć praktyczne zastosowanie w systemach automatycznej translacji mowy i nie tylko.

Ocena jakości tłumaczenia tekstu, będącego wynikiem maszynowego tłumaczenia tekstu, jest nie mniej skomplikowane niż sam proces. Istnieją rozmaite metody oceny tłumaczenia maszynowego. Najprostszy podział metod polega na wyróżnieniu tych dokonywanych przez człowieka oraz tych dokonywanych automatycznie. Najczęściej spotykana jest miara BLEU (ang. *Bilingual Evaluation Understudy*) [10]. Jest ona jedną z pierwszych miar, które wykazały wysoką korelację z oceną jakości dokonywaną przez człowieka. Jest ona obecnie jedną z najpopularniejszych metod oceny jakości. Jej główna idea polega na założeniu, że im bliższe tłumaczenie maszynowe jest profesjonalnemu tłumaczeniu dokonane przez człowieka, tym jest lepsze. Najskuteczniejsze obecnie dostępne systemy automatycznego tłumaczenia tekstu osiągają poziom 60% określony za pomocą miary BLEU.

5. Synteza mowy

W systemach automatycznej translacji mowy synteza mowy odpowiedzialna jest za generowanie wynikowego sygnału mowy na podstawie automatycznie przetłumaczonego tekstu. Synteza mowy jest najbardziej znanym i najbardziej złożonym procesem spośród wszystkich innych rodzajów syntezy dźwięków. Złożoność syntezy mowy w pierwszym rzędzie wynika ze złożoności sygnału mowy na poziomie akustycznym. Drugą przyczyną złożoności tego procesu są wymogi co do określonego poziomu zrozumiałości mowy syntetycznej oraz naturalności jej brzmienia. O ile wymóg naturalności brzmienia ma drugorzędne choć ważne znaczenie, o tyle konieczność zapewnienia odpowiedniej zrozumiałości stanowi podstawowe kryterium umożliwiające konkretne zastosowania syntezy mowy w systemach automatycznej translacji mowy.

6. Usprawnienie procesu automatycznej translacji mowy poprzez zastosowanie algorytmów automatycznej identyfikacji mowy pozwalających na automatyczną segmentację mowy pochodzącej od różnych mówców

Oczekuje się, że prowadzone w ramach grantu nr POIG.01.03.01-24-107/12 prace badawcze w zakresie rozpoznawania i identyfikacji mowy przyczynią się także do usprawnienia procesu automatycznej translacji mowy typu speech-to-speech dzięki zastosowaniu zaawansowanych algorytmów rozpoznawania mowy, pozwalających na skuteczną segmentację sygnału mowy pochodzącej od wielu mówców [11]. Celem automatycznej segmentacji sygnału mowy pochodzącej od różnych mówców jest wydzielenie z sygnału mowy fragmentów zawierających mowę oraz przyporządkowanie ich poszczególnym mówcom. Pożądane jest również przed procesem segmentacji mowy wydzielenie z sygnału fragmentów niepodlegających tłumaczeniu, takich jak muzyka oraz inne dźwięki np. szum otoczenia oraz inne zakłócenia. Konieczność ta wynika z faktu, iż większość systemów rozpoznawania mowy oraz mowy działa poprawnie tylko w sytuacji, gdy rozpoznawana mowa jest wolna od zakłóceń. Poza segmentacją sygnału mowy w wielu zastosowaniach konieczna jest również identyfikacja mówców, czyli przyporządkowanie wydzielonych z sygnału fragmentów mowy do poszczególnych osób, których tożsamość jest znana. Ma to miejsce między innymi w zastosowaniach telekonferencyjnych, gdzie oprócz transkrypcji wypowiedzi istotna jest tożsamość mówiącego. Automatyczna segmentacja sygnału mowy pochodzącej od różnych mówców w systemach automatycznej translacji mowy może działać tylko na podstawie systemów rozpoznawania mowy niezależnych od wypowiedzianego tekstu.

Automatyczne rozpoznawanie mowy (ang. *speaker recognition*) znane w literaturze polskojęzycznej również pod nazwą automatycznego rozpoznawania głosów ARG jest dziedziną wiedzy pokrewną do rozpoznawania mowy. Istotnym elementem odróżniającym te dwa zagadnienia jest fakt, że w rozpoznawaniu mowy istotne jest wydobycie zawartości lingwistycznej z analizowanej wypowiedzi, podczas gdy w rozpoznawaniu mowy najistotniejsze jest wydobycie cech sygnału mowy specyficznych dla danej osoby, które pozwolą na jej późniejsze rozpoznanie.

Automatyczne rozpoznawanie głosów obejmuje automatyczną weryfikację ASV (ang. *Automatic Speaker Verification*), automatyczną identyfikację ASI (ang. *Automatic Speaker Identification*) oraz tak zwaną autentyzację mowy. W procesie weryfikacji użytkownik systemu musi wstępnie zadeklarować swoją tożsamość, wprowadzając swój osobisty numer identyfikacyjny PIN, a następnie jest zobligowany do udzielenia jednej bądź kilku wypowiedzi. System rozpoznawania głosów podejmuje decyzję, czy analizowana wypowiedź została podana przez mówcę o wcześniej deklarowanej tożsamości. Końcowym efektem procesu wery-

fikacji jest potwierdzenie bądź odrzucenie deklarowanej przez użytkownika tożsamości. Podstawą do podjęcia takiej decyzji przez system jest porównanie podobieństwa między wzorcem głosu (zarejestrowanym wcześniej w systemie) a rozpoznawaną wypowiedzią z pewną ustaloną wartością progową. W procesie identyfikacji tożsamość nie jest wstępnie deklarowana, a mówca, którego głos podlega badaniu, może być już uprzednio zarejestrowany w systemie (posiada swój model głosu) bądź jest kimś zupełnie nieznanym dla systemu rozpoznającego. Podczas identyfikacji w zbiorze zamkniętym (ang. *Closed Set Identification*) zakłada się, iż dostęp do systemu mają wyłącznie osoby, których modele głosów zostały wcześniej stworzone. System rozpoznający podejmuje wówczas decyzję typu 1 z N, gdzie N stanowi liczbę zarejestrowanych użytkowników. Gdy założenie to nie jest spełnione, zachodzi identyfikacja w zbiorze otwartym (ang. *Open Set Identification*). Wtedy dodatkowo należy rozważyć, czy podobieństwo cech wypowiedzi nieznanego mówcy do cech jednego z modeli mówców zarejestrowanych w systemie jest na tyle duże, że można podjąć decyzję o zidentyfikowaniu osoby, czy też uznać ją za nienależącą do żadnego z mówców zarejestrowanych w systemie. W drugiej z opisanych sytuacji system może podjąć decyzję o odrzuceniu mówcy bądź też jego zarejestrowaniu. Ostatnią z procedur realizowanych przez systemy rozpoznawania mówców jest autentyzacja, polegająca na ustaleniu, czy wypowiedź należy do jednego z mówców wcześniej zarejestrowanych w systemie czy też nie.

Systemy rozpoznające mówców dzieli się również na zależne (ang. *Text Dependent*) i niezależne od tekstu (ang. *Text Independent*). W systemach automatycznie tłumaczących mowę zastosowanie znajdują systemy niezależne od tekstu, gdyż trudno jest wymagać, aby osoba, którą system musi zidentyfikować, wtrącała w swe wypowiedzi słowa kluczowe. Automatyczna segmentacja sygnału mowy pochodzącej od różnych mówców w systemach automatycznej translacji mowy może działać tylko na podstawie systemów rozpoznawania mówcy niezależnych od wypowiedzianego tekstu.

Dodatkowym elementem usprawniającym proces automatycznej translacji mowy może być zastosowanie algorytmów automatycznego rozpoznawania języka wypowiedzi [12]. Globalizacja i intensywny rozwój telekomunikacji sprawiły, że stopniowemu zacieraniu ulegają granice zarówno polityczne, jak i geograficzne. W rezultacie wzrasta zapotrzebowanie na rozwiązania technologiczne o charakterze uniwersalnym, przełamujące granice językowe, stąd też automatyczna identyfikacja języka na podstawie ograniczonej czasowo wypowiedzi nieznanego osoby staje się w związku z tym istotnym zagadnieniem z punktu widzenia sprawnej translacji mowy.

7. Podsumowanie

W artykule zaprezentowano ideę usprawnienia procesu automatycznej translacji mowy poprzez zastosowanie algorytmów automatycznej identyfikacji mowy, pozwalających na automatyczną segmentację mowy pochodzącej od różnych mówców. Przedstawiono także zasady funkcjonowania typowego systemu automatycznej translacji mowy wykorzystującego: rozpoznawanie mowy, automatyczną translację tekstu oraz syntezę mowy.

Na podstawie przeglądu literatury można odnieść uzasadnione wrażenie, że proces automatycznej translacji mowy jest dość skomplikowany i nie został jeszcze skutecznie rozwiązany oraz wciąż trwają prace nad jego udoskonaleniem. Dlatego też każda propozycja usprawnienia i poprawy skuteczności działania automatycznej translacji mowy wydaje się mieć bardzo duże znaczenie. Liczba projektów badawczych realizowana w dziedzinie automatycznej translacji mowy pokazuje, że ten obszar wiedzy rozwija się w ostatnim czasie niezwykle aktywnie, a realizowane projekty badawcze w tej dziedzinie wspierane są przez rządy wielu krajów na całym świecie [13].

BIBLIOGRAFIA

1. Stuker S., Herrmann T., Kolss M., Niehues J., Wolfel M.: Research Opportunities In Automatic Speech-To-Speech Translation, Potentials. IEEE Volume: 31, Issue: 3, 2012 p. 26-33.
2. Gordon R.G., Jr., Ed.: Ethnologue, Languages of the World, 15th ed. Dallas: SIL International, 2005.
3. Janson T.: Speak—A Short History of Languages. U.K.: Oxford Univ. Press, London 2002.
4. Commission of the European Communities, A new framework strategy for multilingualism, Communication from the Commission to the Council, the European Parliament, the European Economic and Social Committee, and the Committee of the Regions, Nov. 2005.
5. Steinbiss V.: Human language technologies for Europe. Work Comissioned by ITC-irst, Trento, Italy to Accipio Consulting, Aachen, Germany, Apr. 2006.
6. Waibel A., Fügen C.: Spoken language translation—enabling crosslingual human-human communication. IEEE Signal Processing Mag., vol. 25, no. 3, May 2008, p. 70–79.
7. Rabiner L. R., Juang B. H.: Fundamentals of speech recognition. Prentice Hall, 1993.
8. Huang X., Acero A., Hon H.W.: Spoken Language Processing. Englewood Cliffs, NJ: Prentice-Hall, 2001.
9. Koehn P.: Statistical Machine Translation. Cambridge, U.K.: Cambridge Univ. Press, 2009.

10. Muyun Yang, Junguo Zhu, Jufeng Li, Lixin Wang, Haoliang Qi, Sheng Li, Liu Daxin: Extending BLEU Evaluation Method with Linguistic Weight, The 9th International Conference for Young Computer Scientists, ICYCS 2008, p.1683–1688.
11. Dustor A., Kłosowski P.: Biometric Voice Identification Based on Fuzzy Kernel Classifier. Proceedings of 20th International Conference of Computer Networks CN 2013, Lwwek Iski, Poland, June 17-21, 2013, Communications In Computer and Information Science Volume 370, Springer-Verlag Berlin Heidelberg, Germany 2013, p. 456-465.
12. Dustor A., Szwarc P.: Spoken Language Identification Based on GMM Models. International Conference on Signals and Electronic Systems (ICSES), Gliwice, 07-10 September 2010, p. 105–108.
13. Hutchins J.: International Association for Machine Translation compendium of translation software, 2010. [Online]. Available: <http://www.hutchinsweb.me.uk/Compendium.htm>

Wpłynęło do Redakcji 9 kwietnia 2014 r.

Abstract

The article presents selected, effective speech signal processing algorithms and their use in order to improve the automatic speech translation. Automatic speech translation uses natural language processing techniques implemented using algorithms of automatic speech recognition, speaker recognition, automatic text translation and text-to-speech synthesis. It is very possible to improve the process of automatic speech translation by using effective algorithms for automatic segmentation of speech signals based on speaker recognition and language recognition.

Division of Telecommunication, a part of the Institute of Electronics and Faculty of Automatic Control, Electronics and Computer Science Silesian University of Technology, for many years has been specializing in advanced fields of telecommunication engineering. One of them is speech signal processing. The one of many research areas aims to gain new knowledge in the field of the basic phenomena of perception and processing of human speech such as understanding and translation of speech made by a person. The main scientific objective of this research area is development of selected, effective speech signal processing algorithms and their use in order to improve the automatic speech translation. Automatic speech translation system uses natural language processing techniques implemented using algorithms of automatic speech recognition, speaker recognition, automatic translation of text and text-to-speech synthesis. Research hypothesis can be formulated as follows: It is possible to improve

the process of automatic speech translation by using efficient algorithms for automatic segmentation of speech signals coming from different speakers. An expected result of the research project is the development of efficient algorithms which allow to improve the automatic speech translation.

Adresy

Piotr KŁOSOWSKI: Politechnika Śląska, Instytut Elektroniki,
ul. Akademicka 16, 44-100 Gliwice, Polska, piotr.klosowski@polsl.pl

Adam DUSTOR: Politechnika Śląska, Instytut Elektroniki,
ul. Akademicka 16, 44-100 Gliwice, Polska, adam.dustor@polsl.pl

Jacek IZYDORCZYK: Politechnika Śląska, Instytut Elektroniki,
ul. Akademicka 16, 44-100 Gliwice, Polska, jacek.izydorczyk@polsl.pl