STREAM: Prototype Development of a Digital Language Identifier

Rebecca Meehan[1], Manisha Kumari[2], Qiang Guan[2], Sanda Katila[3], Joel Davidson[4] and Nichole Egbert[5]

[1]*School of Information, Kent State University, U.S.A.*
[2]*Department of Computer Science, Kent State University, U.S.A.*
[3]*Visual Communication Design, Kent State University, U.S.A.*
[4]*Akron Children's Hospital, U.S.A.*
[5]*Communication Studies, Kent State University, U.S.A.*

Keywords: Digital Innovation, Health, Language Identifier, First Responders.

Abstract: This paper describes the development of a prototype digital language identifier (STREAM™), designed to help improve health by shortening the time it takes for healthcare professionals and first responders working with non-native speakers to identify a patient's language so that the correct interpreter can be called, enabling the patient to more quickly get the care they need. The innovation was originally designed to address the needs for health care professionals caring for the emergent Bhutanese and Nepali community of newcomers in Akron, Ohio, USA. Language access support through in person and electronic interpreters continues to improve, however, there remains a need to quickly identify spoken language at points of entry or in emergent situations. We developed and tested a digital prototype solution (Smart Translation Enabling and Aiding Multicultural populations, aka "STREAM™" tool, patent pending) based on the Nepali language that can be later extended to identify multiple languages. Prototype testing of STREAM™ showed that although the model predicted the correct language better than chance, accuracy needs to be improved. Next steps include refining the model to identify spoken language in a shorter amount of time, adding other languages to the model, and user testing among medical and emergency services professionals.

1 OBJECTIVES

This paper describes the pilot development of a prototype digital language identifier ("STREAM™"). The objectives of the paper include a report of the pilot development process and testing: (1) purpose of the prototype, (2) status of the prototype, (3) navigation flow of the prototype, (4) technical specifications, (5) testing and results demonstrating successful identification of the Nepali spoken language, and (6) next steps.

2 PURPOSE OF PROTOTYPE

The purpose of the prototype, "STREAM™", is to help first responders and health care professionals when they encounter community members in needof health care or emergency services, and whose spoken language is unclear. This situation often arises within cities and regions that have a new or growing population of refugees or newcomers who may not speak the regional language, and where first responder and health care professionals are unable to discern the language of these newcomers or refugees. Currently, first responders and health care professionals may employ tactics such as displaying language identification cards for the patient to point to their spoken language. They may also rely on their best guess and initiate interpreters for the language they suspect the person is speaking. However, problems arise when the wrong language is identified, the speaker cannot read, the card/poster is not present or used, or the language is not listed on the resource. STREAM™ (**S**mart **Tr**anslation **E**nabling and **A**iding **M**ulti-cultural populations) is a digital language identifier used to identify a patient's spoken language, so that the appropriate interpreter can be engaged in a health care (arriving in person or calling for an appointment) or emergency situation. STREAM™ will use spoken words from the patient, process

Meehan, R., Kumari, M., Guan, Q., Katila, S., Davidson, J. and Egbert, N.
STREAM: Prototype Development of a Digital Language Identifier.
DOI: 10.5220/0010893200003123
In *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022) - Volume 5: HEALTHINF*, pages 664-670
ISBN: 978-989-758-552-4; ISSN: 2184-4305

audio through an artificial intelligence (AI) algorithm and then display the language results with words, visual symbols and audio clues. Identifying a patient's language fasterwill assist service providers in communicating vital health or emergency information, expediting appropriate care, promoting better health outcomes and improved satisfaction of staff and patients. STREAM™ (patent pending) aims to improve refugee health by reducing the time it takes to connect with the right interpreter, facilitating care delivery to the person in need.

2.1 Background

STREAM™ was developed to address the concerns of local health care leaders in the Akron, Ohio area who witnessed how difficult it became for health care professionals and first responders toquickly identify the language of a group of newcomers or refugees. Akron, Ohio's largest group of newcomers is ethnically Bhutanese and Nepali speaking. Since 2008, approximately 80,000 Nepali speaking Bhutanese adults and children have resettled in the United States (Roka, 2017). As of 2020, estimates of the Bhutanese community in our local area of Akron, Ohio within Summit County, range between 4,000 and 5,000 people (Case Western Reserve University, 2020). In 2013 Akron's foreign-born population held $137million in spending power and between 2000-2013 has increased the total housing value in Summit County by $207 million (Partnership for a New American Economy & Knight Foundation, 2016; Vigdor et al., 2013). The community is growing and showing signs of thriving as evidenced by Bhutanese – Nepali specialty food stores, religious services, traditional music and community festivals (Chmura Economics & Analytics, 2017).

2.2 Health Care

Health care professionals can encounter patients over the phone, or in person, both in emergency/acute or non-emergency situations, for example setting up an appointment or asking for information. Whether a patient calls on the phone or arrives in person, a nurse or receptionist may notice the person struggling with English, ask the person what language they speak, or make a best guess, then initiate interpreter services or choose toproceed without connecting with an appropriate and available interpreter. In order to help address the health concerns of the patient efficiently, before an interpreter is called, the health care worker needs to know the language of the speaker. First responders encounter these issues as well.

2.3 First Responders

First responders include police, emergency services/ 911 dispatchers, EMS, and firefighters. Emergency services/ 911 dispatchers receive thousands of calls each year, including calls where a non-English speaker is trying to obtain first responder help. If the 911 dispatcher cannot understand the language, they often leverage an Emergency Interpreter Service (Zimmermann, 1996) to interact with the caller. Despite the availability of this resource, the 911 dispatcher still needs to know what language the speaker is using so that the appropriate interpreter can be utilized asquickly as possible. The moments waiting for the appropriate translator to join the call have been characterized as "tense," especially when the dispatcher does not recognize the language spoken (VOA, 2014). Because time is critical to addressing emergency situations, any mechanismto bring a successful resolution to the call is valuable, to address the emergency appropriately. Other first responders, including the police, may seek to understand someone's spoken language by using a protocol for helping someone with limited English proficiency (LEP) (Under VI of the Civil Rights Act of 1964 (42 U.S.C. §2000d et seq.). While the process can vary, the protocol involves asking for a person's language identification card, or providing a list of languages that a person could look through and point to their spoken language. Other officers may troubleshoot by asking bi-lingual colleagues to assist them. Still, the language of the patient or person in need may still be unknown.

2.4 Current Tools for Language Identification

The U.S. Department of Health and Human Services (DHHS) and the Department of Homeland Security (DHS) developed specific toolsto help someone with LEP to communicate their language in order to seek help from first responders or health care professionals. These include Language Identification cards and the "I Speak" resources. A person who does not speak English can carry a language identification card to show in case of emergency or in everyday activities. The "I Speak" resource includes both poster form and a booklet listing 70 languages written in those languages. Additionally, health care facilities often display a poster at the check in area of a health care office, including a number of languages, and allowing a non-English speaking patient to identify the language they speak. All of these visual tools, however, are not helpful if the person or patient is not literate, nor if they do not have the visual cue (poster, booklet or identification card) with them at the time of the health encounter or emergency. A digital solution, like STREAM™, with both visual and audio cues, accessible from any mobile device with access to the internet, would allow for greater flexibility in identifying a person's language.

3 STATUS OF PROTOTYPE: STREAM™

STREAM™ (patent pending) is in the early stages of development and pilot testing. STREAM™ is a digital solution, accessed on a smartphone, tablet or computer, that "listens" to phrases spoken in the speaker's language for about 20 seconds and then, through using an artificial intelligence (AI) and machine learning (ML) multi-language classifier algorithm, visually displays to the health care professional or first responder and the speaker (patient) what language is being spoken. Once there is a confirmation between patient and health care professional or first responder, the correct interpreter can be called to help with the emergency or the medical appointment. This process can also facilitate an initial set of pre-programmed instructions or phrases for the patient as to where they should go or do next.

The STREAM™ prototype was built to address current gaps in the Akron, Ohio community for identifying Nepali language speakers, Akron's largest newcomer group. This manuscript does not involve human subjects and does not require IRB approval. There are about 7,117 languages which are spoken in the world (Eberhard et al., 2020). While the STREAM™ tooluses a multi-language classifier algorithm and natural language samples from over 176 languages,for the purposes of prototype testing, the tool is currently built to identify one language of lesser diffusion (National Council on Interpreting in Health Care, n.d.): Nepali. In future development, STREAM™ will be trained to identify other spoken languages to meet the needs of diverse populations. This prototype demonstrates it is possible to identify spoken language for those languages of lesser diffusion. The intended primary end-users of STREAM™ include health care professionals and first responders (e.g. police, 911 dispatchers, firefighters, EMS, etc.). STREAM™ is not built as an app, but instead, it is designed in a "software as a service" (SAAS) model, a cloud-based system, allowing for seamless updates without requiring the end user to reload or update their app. End users will bookmark the URL website for STREAM™ on their mobile phone, tablet or computer and keep it open or easily accessible to begin an interaction whenever needed.

4 NAVIGATION WORKFLOWOF THE PROTOTYPE

A typical use case and navigation workflow for STREAM™ is pictured in Figure 1 and outlined in the steps below:

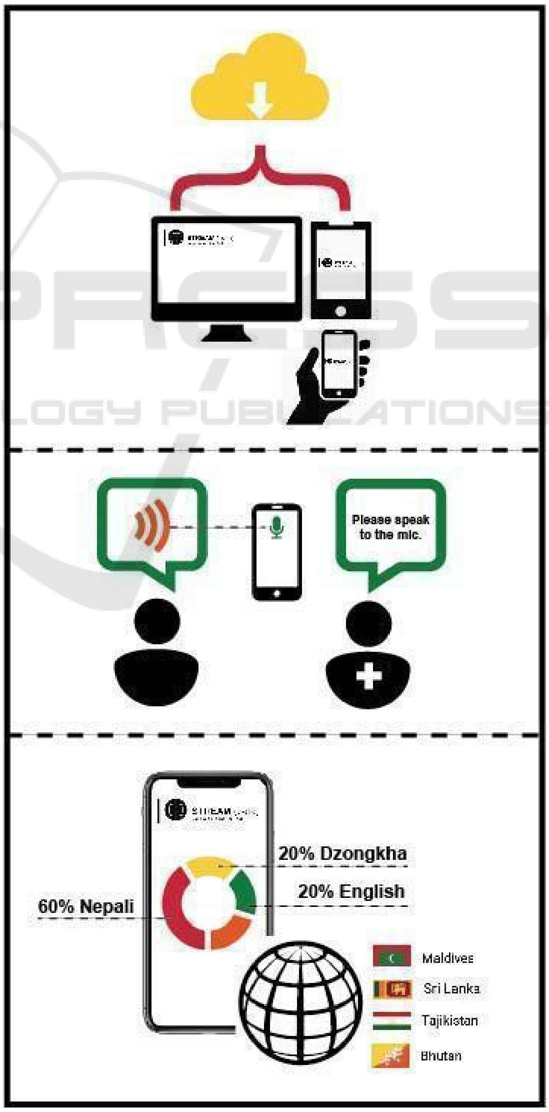


Figure 1: Process for Using Stream.

1. Health care professional or first responder (end-user) opens the STREAM™ site.
2. When the patient is ready, the end-user clicks on the “Identify Language” button, and has the patient speak into the microphone of the user’s device (e.g. mobile phone, computer, tablet, etc.) that will “listen” to the person speaking. Visual cues will prompt users to speak into the system.
3. After about 20 seconds, the STREAM™ tool determines results by displaying a list of the most likely languages, with the most likely spoken language displayed prominently alongwith other visual cues.
4. STREAM™ then prompts by displaying the phrase: “Were you speaking X language?” along with visual symbols and auditory cues of the language (e.g. flags, maps, etc.)

- If yes, the user (health care professional or first responder) then initiates connection with the appropriate interpreter or follows their next intake protocol steps.
- If no, the user can point to the 4 other likely languages on display to see if one of them is the correct language.

5 TECHNICAL SPECIFICATIONS

The STREAM™ prototype is based on a software stack that includes two major components: STREAM™ Engine (SE) and STREAM™ Web Service (SWS), shown in Figures 2 and 3, respectively. The SE (Fig. 2) is the heart of this prototype, as it addresses the tasks of audio-data pre-processing, model training and model deployment.

Our language prediction algorithm is trainedon a blended speech audio dataset including over 90,000 audio clips from 176 different languages. Audio clips are preprocessed in the same length, same file format as WAV files and converted into the numeric array after extracting the features using Mel Frequency Cepstral Coefficients (MFCCs) feature extraction method (Bhandari, 2020; Koolagudi et al., 2012; Shmueli, 2019). The noise removal filters are also applied to enhance the prediction precision. The project utilizes a multi-classification model to show multiple possible results of spoken language, with language with the highest prediction at the top of the list, followed by four others. Although the model started with using 176 + 1(Nepali) languages, the accuracy of the prototype was lower than desired (around 60%), the final iteration of the prototype demo used 12 languages and added new sample data from YouTube for each of the 12 languages. The final datasets for the model have some variations (different speakers, male and female voices). We gave more preferences to Nepali by adding more sample data compared to other 11 languages in the final model. You can see these 12 languages in the F1 score picture (see Table 2).

After audio data preprocessing, we applied multiple machine learning algorithms including eXtreme Gradient Boosting (XGBoost), Support Vector Machine (SVM), Long Short-Term Memory (LSTM) etc. to develop the classification models and used ensemble techniques to combine the models to increase the accuracy. We also performed hyper parameter tuning and audio data augmentation to optimize the model performance. The accuracy of the model is identified using Confusion Matrix and Area Under the Receiver Operating Characteristic Curve (AUC-ROC) score. For the final solution, to increase the model performance we have compromised the model accuracy very slightly and use only the XGBoost

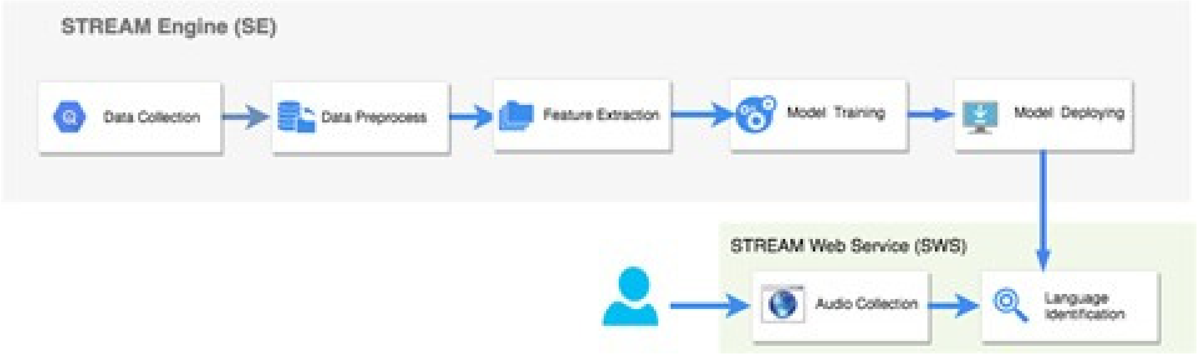


Figure 2: STREAM™ System Architecture / STREAM Engine (SE).

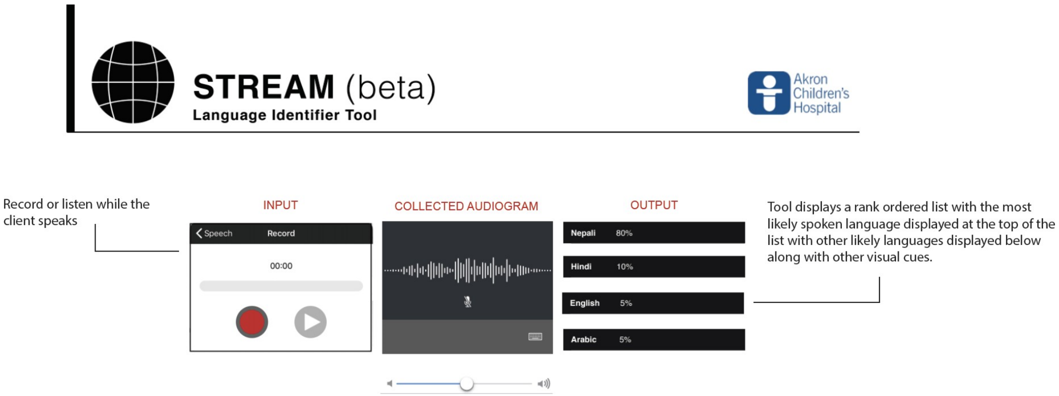

STREAM (beta)
Language Identifier Tool
Akron Children's Hospital
Record or listen while the client speaks
INPUT
Speech
Record
00:00
COLLECTED AUDIOGRAM
OUTPUT
Nepali 80%
Hindi 10%
English 5%
Arabic 5%
Tool displays a rank ordered list with the most likely spoken language displayed at the top of the list with other likely languages displayed below along with other visual cues.

Figure 3: Prototype of STREAM™ Web Service (SWS).

multiclass algorithm to train the language prediction model with fine-tuned parameters for the Nepali language specifically as the target language for the prototype. After the model is evaluated, if it satisfies the required prediction accuracy, it will be deployed to the STREAM™ Web Service (SWS) host for the prototype (shown in Fig 3).

SWS provides human-computer interactive service, which allows the end users to input a short sample of speaking (20 seconds currently). SWS will automatically process the audio data and parallelly feed the model(s) with the input clips of the audio file. The SWS user interface will be tested and modified in upcoming research and development. (See future research section below.)

6 TESTING AND RESULTS

In training the prototype algorithm, the whole dataset is split into a training dataset and a testing dataset. An independent validation dataset was collected from YouTube (with Nepali Speakers). In order to avoid an overfitting problem, we utilize a k-folder mechanism to split the dataset into k folders and each time one folder is selected as the validation dataset and the rest of the folders areused for training the model. After k iterations, we average the accuracy. After training the model, we validate the model and use the independent dataset. While used on site, the prototype we built requires the audio inputs from speakers with thelength of 20-30 seconds. The audio captured willbe further duplicated with multiple copies (in ourcase, we use five copies) to feed different identifiermodels. The final language identification will be determined after summarizing the predictions from different models.

6.1 Prediction Results

In order to evaluate the results of the prototype predicting or identifying the correct spoken language (Nepali), we use precision, recall, and F1-score to illustrate the performance of our model detection capability. We show the truth and prediction condition in the following table (Table 1):

Table 1: Prediction Results Format.

Total Samples		Prediction	
		Negative	Positive
Truth Condition	Negative	True Negative	False Positive
	Positive	False Negative	True Positive

Precision, recall and f1-score are defined accordingly:

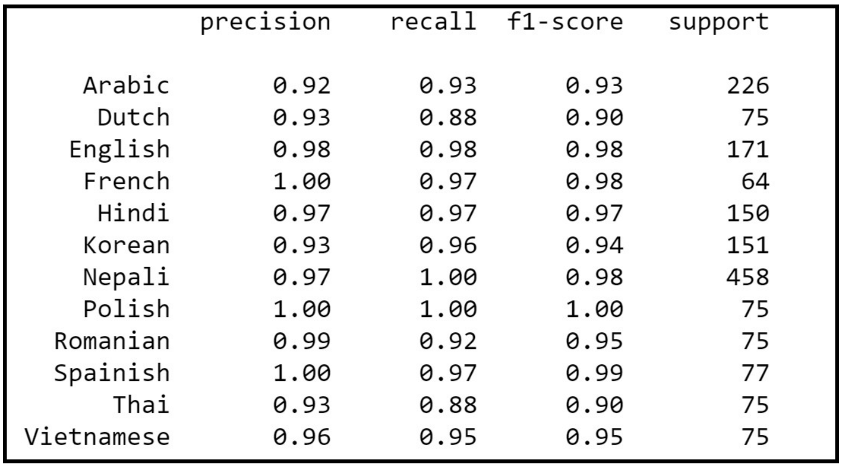
$$\text{Precision} = \text{(True Positive)/(True Positive + False Positive)} \quad (1)$$

$$\text{Recall} = \text{(True Positive)/(True Positive+ False Negative)} \quad (2)$$

$$\text{F1} = \text{2x(Precision Recall)/(Precision+Recall)} \quad (3)$$

Precision explains how accurate the model is out of the predicted positives. Precision value will be between 0 and 1 (a higher value means the model is more accurate). Recall addresses if the model has found out all positives with value between 0 and 1 (a higher value means the model is more robust to the false negatives). F1-score is based on precision and recall to balance between precision and recall with its best value at 1 and worst at 0.

Table 2: XGBoost Multi-Classification Report.



```
             precision    recall  f1-score   support

      Arabic       0.92      0.93      0.93       226
       Dutch       0.93      0.88      0.90        75
     English       0.98      0.98      0.98       171
      French       1.00      0.97      0.98        64
       Hindi       0.97      0.97      0.97       150
      Korean       0.93      0.96      0.94       151
      Nepali       0.97      1.00      0.98       458
      Polish       1.00      1.00      1.00        75
    Romanian       0.99      0.92      0.95        75
    Spainish       1.00      0.97      0.99        77
        Thai       0.93      0.88      0.90        75
  Vietnamese       0.96      0.95      0.95        75
```

The prototype was successful in meeting the objective of identifying the Nepali spoken language. This is demonstrated by the final XGBoost model for identifying Nepali showing 97% of precision and 100% of recall with F1-score as 0.98. The F1 score for all different languages was more than 90% shown in the table above (Table 2).

7 NEXT STEPS

The language identifier, STREAM™, is currently in a prototype phase, using AI to generate a list of the top 5 potential languages that a patient is speaking. Next steps in the prototype development development include further refining the tool to a) improve accuracy of results; b) accurately identify the Nepali language (Nepali) in a shorter amount of time; c) training the tool to identify additional languages and dialects with greater accuracy, and d) conducting user testing for feedback on improving workflow and user experience of the tool. The benefit of using the AI algorithm is that it will continue to improve in accuracy as it is trained and used, as it continues to learn. Further development will use not only more languages and dialects, but also greater volume of samples, so as to improve accuracy of results. Moreover, development will incorporate contextual application differences for in person, emergency and telephone uses.

The best opportunity for developing the prototype to be most effective is by evaluating the use of the device in the context of applied user scenarios. To that end, user testing of the tool will be conducted to examine the process for identifying unknown spoken languages in multiple settings (e.g. 911 dispatch centers, community calls by police and firefighters, emergency departments, and health care centers) and in multiple modalities (e.g. telephone and in person). Specifically, user testing will look at how the end-user accesses the STREAM™ system, how they prompt patients to speak into the system, and the reaction, results and follow up questions when the results are made known. Future research will examine the influence of STREAM™ on reduced wait time for patients; reduced costs for emergency and health care systems; patient satisfaction; and satisfaction of both health care professionals and first responders.

REFERENCES

Bhandari, A. (2020). *AUC-ROC Curve in Machine Learning Clearly Explained*. https://www.analyticsvidhya.com/blog/2020/06/auc-roc-curve-machine-learning/

Case Western Reserve University. (2020). *Encyclopedia of Cleveland History*. https://case.edu/ech/

Chmura Economics & Analytics. (2017). *Economic Impact of Refugees in the Cleveland Area, Calendar Year 2016*. Refugee Services Collaborative of Greater Cleveland. http://rsccleveland.org/wp-content/uploads/2016/12/ EconomicImpactStudy2017FINAL.pdf

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (2020). Ethnologue: Languages of the world. 23rd edn. Dallas: SIL International. *Online: https://www. Ethnologue. Com*.

Koolagudi, S. G., Rastogi, D., & Rao, K. S. (2012). Identification of language using mel-frequency cepstral coefficients (MFCC). *Procedia Engineering*, *38*. https://doi.org/10.1016/j.proeng.2012.06.392

National Council on Interpreting in Health Care. (n.d.). *Languages of Limited Diffusione*. Retrieved September

6, 2020, from https://www.ncihc.org/ languages-of-limited-diffusion

Partnership for a New American Economy, & Knight Foundation. (2016). *Welcome to Akron: How Immigrants and Refugees Are Contributing To Akron's Economic Growth*. https://www.newamericaneconomy.org/wp-content/uploads/2016/06/Welcome-to-Akron_Partnership-for-a-New-American-Economy_June-2016.pdf

Roka, K. (2017). Adjusting to the New World: A Study of Bhutanese Refugees' Adaptation in the US. *Journal of Sociology and Social Work*, *5*(2). https://doi.org/10.15640/jssw.v5n2a11

Shmueli, B. (2019). *Multi-Class Metrics Made Simple, Part II: the F1-score*. https://towardsdatascience.com/multi-class-metrics-made-simple-part-ii-the-f1-score-ebe8b2c2ca1

Vigdor, J. L., Calcagni, G., & Valvekhar, M. (2013). Immigration and the Revival of American Cities:From Preserving Manufacturing Jobs to Strengthening the Housing Market. [online] Washington, DC: Partnership for a New American Economy. *Available at: [Accessed 24 March 2017]. White.*

VOA. (2014). *It's an Emergency in Any Language*. https://learningenglish.voanews.com/a/emergency-911-multi-language/2429774.html

Zimmermann, P. G. (1996). Use of interpreters in the emergency department. *Journal of Emergency Nursing: JEN: Official Publication of the Emergency Department Nurses Association*, *22*(3). https://doi.org/10.1016/S0099-1767(96)80114-0