INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Exploring the Correlation of Pitch Accents and Semantic Slots for Spoken Language Understanding

Sabrina Stehwien, Ngoc Thang Vu

University of Stuttgart, Germany
{sabrina.stehwien,thang.vu}@ims.uni-stuttgart.de

Abstract

We investigate the correlation between pitch accents and semantic slots in human-machine speech. Using an automatic pitch accent detector on the ATIS corpus, we find that most words labelled with semantic slots also carry a pitch accent. Most of the pitch accented words that are not associated with a semantic label are still meaningful, pointing towards the speaker's intention. Our findings show that prosody constitutes a relevant and useful resource for spoken language understanding, especially considering the fact that our pitch accent detector does not require any kind of manual transcriptions during testing time.
Index Terms: spoken language understanding, speech recognition, pitch accent detection, semantic slots, ATIS

1. Introduction

The connection between various prosodic characteristics of speech and the meaning of discourse in languages like English, German and Dutch is a well-researched field. For example, prosody can interact with syntax [1, 2, 3] and pitch accents can also mark different types of information status (given vs. new information) or categories of information structure (focus, contrastive topics) [4], e.g. *What did you see? - I saw a BIRD* (accented and new information), *I want to go to AUSTRALIA* (focus is accented).

Since prosody provides essential discourse information that is available only from spoken language, there has been a significant amount of research towards its use for Automatic Speech Recognition (ASR) [5, 6, 7, 8, 9] and Spoken Language Understanding (SLU) [10, 11, 12, 13] as well as its impact on ASR errors [14, 15]. On a similar motivational basis as our work, Shriberg and Stolcke [13] use prosodic modelling to improve ASR and several subtasks of SLU. The authors also focus on detecting information units such as topics [12] without requiring prosodically pre-labelled data. Starting from time-aligned speech data, they extracted features that capture e.g. duration, fundamental frequency and voice quality and obtained promising results for automatic speech segmentation. Motivated by fact that items that have been introduced in the discourse are often deaccented, Rösiger and Riester [16] found that the presence or absence of a pitch accent is helpful in improving automatic coreference resolution.

Typical approaches to understanding speech model ASR and the succeeding SLU and NLP tasks disjointly and thus optimize them separately. ASR is optimized in order to create output transcriptions that are as precise as possible, ideally recognizing every uttered word correctly. When uncovering semantic content for SLU, it may be useful to already have a notion of where in the text the most important information is located. Prosodic information such as pitch accents can point towards the most salient information in a text before any deeper linguistic analysis is applied. The main tasks that SLU comprises – domain detection, intent detection and slot filling [17] – may benefit from access to such a resource. For example, content words, being frequently pitch accented, may help determine the domain of speech data, e.g. *I'd LIKE to book a FLIGHT*. Phrases with typical intonation patterns like *SHOW me*, *I WANT to KNOW* and question rises can unveil the speaker's intent.

We investigate the correlation between pitch accents and words that are annotated with semantic labels, also referred to as semantic slots, on the standard corpus for SLU research, the Airline Travel Information System (ATIS) corpus [18]. These semantic labels are assigned to words that express important content, such as location, date and time. Our goal is to determine whether pitch accents can help in localizing this type of information from raw speech data. To address this notion, we require an automatic pitch accent detecting setup that does not rely on manual transcriptions of speech. We can use this detector to find pitch accents in the ATIS dataset and, using the knowledge of where the semantic slots are located, examine where and on what words accents and slots co-occur.

2. Pitch Accent Detection

In this section, we present a method to acquire the locations of pitch accents from speech files and time-aligned text after automatically obtaining the transcriptions from an ASR system. This way we can bypass the manual transcription process during testing time. Manual labelling is only required to train the model, which we describe in the following.

2.1. Data

We train a pitch accent detector on a subset of the Boston Radio News Corpus [19] that is partially labelled with prosodic events (ToBI accent and boundary types). In this work we refer to this subset as the Boston Prosody Subset (BPS). This subset encompasses 220 speech files by 5 speakers, 3 female and 2 male, and consists of in total around 1 hour and 21 minutes of speech. We consider the binary case (pitch accent or none) and group all pitch accent types together as one class. Statistics per speaker are shown in Table 1.

Table 1: *Statistics of the Boston Prosody Subset.*

Speaker	f1a	f2b	f3a	m1b	m2b
# Files	42	100	16	30	32
# Words	1,865	6,994	1,142	1,907	1,928
# Accents	1,126	3,978	609	1,022	1,153
Accent Ratio	0.60	0.56	0.53	0.54	0.60

http://dx.doi.org/10.21437/Interspeech.2016-511

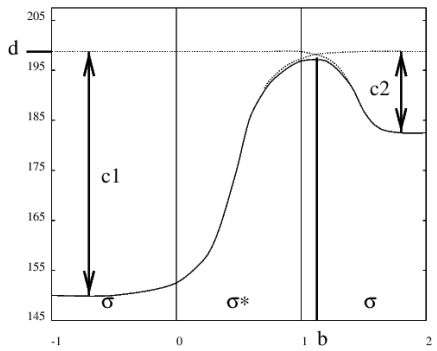

Figure 1: *Parametrized Intonation Events, taken from Möhler and Conkie [27]. a1 and a2 describe the steepness of the rising and falling sigmoid representing the F0 contour, d and b represent the height and temporal alignment of the peak, and c1 and c2 describe the amplitude of the rising and falling sigmoid.*

2.2. Model

There has been a substantial amount of research towards the detection and classification of pitch accents using various methods and modelling approaches [20, 21, 22, 23] while others have concentrated on feature design [24, 25, 26]. Approaches using only acoustic features have reached around 79% on average on the Boston Radio News Corpus. An overview of previous research and performance is given in [25].

Our model uses acoustic features extracted for single syllables inspired by Schweitzer [24]. The 23 features were shown to be helpful in categorizing pitch accents and largely describe the fundamental frequency contour. These include z-scores (speaker normalizations) of the six PaIntE parameters [27] (a1, a2, b, c1, c2, d) shown in Figure 1 and further amplitude descriptions (maximum of c1 and c2, difference between c1 and c2) as well as the duration of the syllable nucleus. The rest of the features comprise c1, c2, d and the nuclei durations for the two preceding and following syllables.

We use a random forest model trained using WEKA [28] to create a binary classifier that detects whether a stressed syllable carries a pitch accent or not. By training 5 separate models, each leaving one different speaker out for training and then testing on the left-out speaker, we obtain 74.4% accuracy on average. Table 2 shows the per-speaker accuracy of the model. Considering the fact that we built the model solely from F0 and nuclei duration features without any further acoustic or lexical information, the performance is decent. Ranging from around 73% to 77% accuracy, the model proves quite stable across the various speakers.

Table 2: *Accuracy of Pitch Accent Detection per speaker on the Boston Prosody Subset considering stressed syllables*

Speaker	Accuracy (%)
f1a	73.1
f2b	74.7
f3a	76.7
m1a	73.7
m2b	73.8

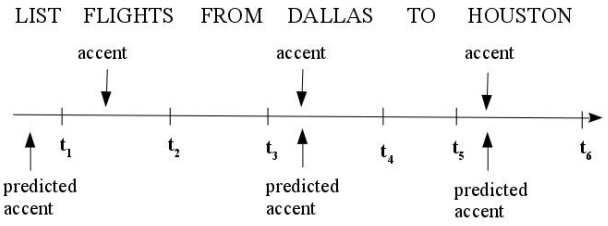

Figure 2: *Word-level evaluation. For each word in the reference transcription, we count a true positive hit if there is a pitch accent during the complete word duration in both the reference (second row) and the predicted labels*

2.3. Pitch Accent Detection from Audio Only

The detection of prosodic events like pitch accents requires the speech files to be time-aligned with the respective phone-, syllable- and word-level transcriptions. In a classic SLU setting, however, the only input is the speech recording, which is automatically transcribed using ASR. We simulate a setup where we have no transcriptions available, and intend to detect pitch accents using the audio file only. We start by automatically recognizing the transcriptions. The acoustic model for our ASR system is trained on the training set of the Wall Street Journal (WSJ) corpus [29] using a neural network model setup (nnet2) in the Kaldi ASR toolkit [30]. As a language model, we interpolate a WSJ Kneser-Ney 4-gram and a BPS bigram model (weights = 0.5) with which we obtain a perplexity of 205 on the BPS. After recognizing the data, we measure a word error rate (WER) of 27.4 %. Finally, we create phone, syllable and word alignments for the recognized transcription. Using our speaker-independent pitch accent detection models, we predict the time points of pitch accents using (1) the recognized output and (2) the reference transcriptions for each speaker individually.

2.4. Results

Since we are interested in pitch accents that lie on certain words, not single syllables, we compare the results at the word level with the reference accent labels. We use the following evaluation method to roughly determine how well our detector performs on the word level illustrated in Figure 2: We consider the time intervals from t_{n-1} to t_n in which each word w_n lies. If a reference accent (second row) is located at a timestamp within this interval, then this word is taken as accented. If a predicted accent is found within the same interval, then we count a true positive hit. It is possible to have several accents on one word (e.g. words with several stressed syllables like *transportation*), in which case this method only counts one accent per word.

The results for the first scenario in which we use the reference transcription are shown in Table 3. The F1-scores obtained on the male speakers (m1b, m2b) are considerably lower (62.5%) than those obtained on the female speakers (>70%). Table 4 contains the results for the same procedure using the recognized transcriptions. Here we observe a much lower variance across speakers. The averages of these measures over all speakers are listed in Table 5. The precision is higher after using the reference transcriptions, but lower using the recognized transcriptions. Considering the the average F1-score lies around 75% for both versions, we conclude that manual transcriptions are not necessary for localizing pitch accents in a speech recording. This finding suggests that pitch accent detection can be readily integrated into SLU tasks.

Table 3: *Word-level accuracy of pitch accent detection per speaker on the BPS using reference transcriptions*

Speaker	f1a	f2b	f3a	m1b	m2b
# Predicted	1,005	4,459	525	665	815
# True positives	817	3,348	406	527	655
Precision (%)	81.3	75.1	77.3	79.2	79.2
Recall (%)	72.6	84.2	66.7	51.6	51.6
F1-Score (%)	76.7	79.4	71.6	62.5	62.5

Table 4: *Word-level accuracy of pitch accent detection per speaker on the BPS using recognized transcriptions*

Speaker	f1a	f2b	f3a	m1b	m2b
# Predicted	1,146	4,527	690	1,027	1,130
# True positives	855	3,195	468	749	867
Precision (%)	74.6	70.6	67.8	72.9	76.7
Recall (%)	75.9	80.3	76.8	73.3	75.2
F1-Score (%)	75.3	75.1	72.1	73.1	76.0

3. Correlation of Pitch Accents with Semantic Slots in ATIS

3.1. The ATIS Corpus

In this experiment, we work on the ATIS corpus; it contains both speech data and corresponding text files that are annotated with semantic labels. The speech files consist of single utterances by speakers requesting flight information from a dialog system, for example *Show flights from Burbank to Milwaukee for today*. This dataset is used for SLU tasks such as slot filling, in which semantic roles like *departure*, *destination* and *departure date* are assigned to the respective terms in each corpus (in this example, *Burbank*, *Milwaukee* and *today*). These slots describe the key query terms, which can be considered as the most important information in this setting. For this reason, we expect the speakers to put special prosodic emphasis on these words in the form of pitch accents.

3.2. Experimental Setup

We use a subset of the ATIS corpus (607 files of the ATIS3 test set) that is annotated with semantic labels as shown in the top part of Figure 3 and use a Kaldi ASR model (triphone model with LDA features) trained on the ATIS2 and ATIS3 Class-A (context-independent) training data to recognize the test set (WER 11.7%) and time-align it to the audio files. We use our pitch accent detector trained on all speakers of the BPS to predict the locations of pitch accents on this dataset.

3.3. Results

In order to measure the correlation of the predicted pitch accents with slots, we use a similar evaluation procedure as in section 2.4: We consider pitch accents located in the time interval of relevant words (words annotated with semantic slots), as in Figure 3. We are interested in how many slots co-occur with predicted pitch accents and in what cases pitch accents occur on non-slot words. We consider the time intervals from t_{n-1} to t_n in which each word w_n lies if it is annotated with as a slot (here: *DALLAS, HOUSTON*). If a predicted accent is found within the same interval, then we count this as a slot that carries an accent. We also count predicted accents that are found within the intervals of words that are not labelled with a slot (e.g. *FLIGHTS* in this figure) and, vice versa, slot words that do not have an accent (not pictured). Table 6 shows the results.

Table 5: *Average word-level accuracies of pitch accent detection using automatic and reference transcriptions of the BPS. 220 files, 13836 words, 7888 accents in reference*

	reference	automatic
# Predicted	7,460	8,520
# True positives	5,753	6,134
Precision (%)	77.0	72.0
Recall (%)	72.9	77.8
F1-Score (%)	74.9	74.8

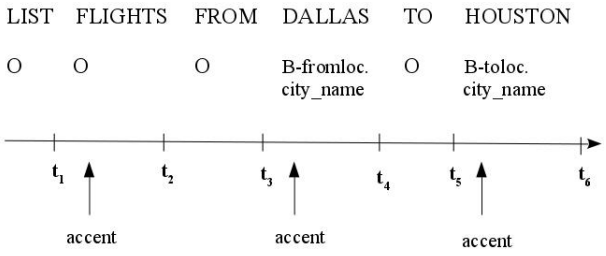

Figure 3: *Estimation of correlation between words and slots. We consider how many times a predicted pitch accent lies within the time interval of a word that is annotated with a slot label*

Table 6: *Frequency of predicted pitch accents in ATIS*

# Files	607
# Words	6,099
# Slots	2,452
# Predicted accents	3,428
# Pred. accents on slots	2,218
# Pred. accents on non-slots	1,210
Slots with pred. accent	90.5 %

Our pitch accent detector predicts around 3,400 accented words of which 2,218 are annotated as slots. This result shows that we can cover 90.5% of all the slots in the dataset using the pitch accent detector, which means most of the information required for slot filling can be localized by finding pitch accents. Since we are also interested in what other information we can localize in this manner, we consider the 1,210 pitch accented words that are not associated with slots in the next analysis. Table 7 contains a list of the non-slot words that most frequently received a predicted pitch accent. Words that do not have semantic labels but are frequently accented are e.g. *list, what, which, please, show* and *need*. These are question words and imperatives that show the speaker's intention, namely requesting the program to provide information in flights. Among these words are also *flight* and *flights*, which are indicative of domain. Frequent non-slot words that are not often pitch accented are function words like *to, from, and, the*, which is typical of intonation patterns in English. The word *flight* and its plural form *flights* occur quite often in ATIS and are accented around 60% of the time. These are important content words, however since the semantic content of this corpus revolves around airtravel information, it may often be regarded by the speakers as *given* information, which is usually deaccented [2].

In summary, our results provide evidence that the localization of pitch accents can directly be used to find important information from speech data which may be helpful for SLU.

Table 7: *Most frequent accented non-slot words in ATIS*

Word	accented	frequency	percentage
FLIGHTS	179	313	57.2
LIST	137	157	87.3
FLIGHT	93	136	68.4
TO	87	504	17.3
ON	83	160	51.9
WHAT	79	87	90.8
I	78	106	73.6
FROM	78	448	17.4
ME	76	102	74.5
SHOW	63	83	75.9
NEED	39	50	78.0
ALL	37	48	77.1
WHICH	32	34	94.1
AND	31	73	42.5
PLEASE	31	33	93.9
THE	29	184	15.8

4. Agreement with Human Labelling

4.1. Pitch Accent Detection Performance on ATIS

In order to determine how well the pitch accent detector performs on ATIS, we let a human labeller annotate 50 files with pitch accents. It is, however, a well-known issue in the prosodic community that the inter-annotator agreement remains around 80% [31]. Therefore, it is difficult to determine absolute performance and we instead estimate the agreement. Table 8 shows that in total around 230 words were found to carry pitch accents by both the predictor and the labeller. Specifically, they agree in 173 cases, which indicates that our detector is reasonably accurate on this dataset. This also shows that the results reported in the previous section can be considered reliable.

Table 8: *Performance of the pitch accent detector, evaluated using 50 manually labelled files of the ATIS subset*

# Files	50
# Words	514
# Slots	201
# Human-labelled accents	235
# Words with predicted accents	234
Agreement: # words	173

4.2. Coverage of Semantic Slots using Pitch Accents

We compare the correlation between slots and both the human-annotated and the automatically predicted pitch accents. Tables 9 and 10 list the details of this analysis. Around 74% of slots are judged by the human labeller to bear a pitch accent, while around 82% were predicted as accented by the automatic detector. These numbers indicate that in this specific subset, the predicted pitch accents correlate more with slots than those created by the human labeller. Based on these results, it may be concluded that the detector is more suitable for localizing slots.

Table 9: *Correlation between slots and human-annotated pitch accents in 50 ATIS files*

# Human-labelled accents	235
# Accents on slots	149
# Accents on non-slots	86
# Slots with no accent	52
# Slots with accent	74.1 %

Table 10: *Correlation between slots and automatically predicted accents in 50 ATIS files*

# Predicted accents	234
# Accents on slots	164
# Accents on non-slots	70
# Slots with no accent	37
# Slots with accent	81.6 %

5. Discussion and Future Work

This study presents promising results on ATIS as the well-researched benchmark corpus for SLU. Since it consists of human-to-machine speech recorded in a laboratory setting [18], its naturalness may be questioned. Further experiments on other datasets are necessary to draw more general conclusions.

State-of-the-art SLU methods reach accuracies up to around 95.6% on the slot filling task [32, 33]. In contrast, the performance of slot filling has been shown to drop considerably on recognized text [34]. In such cases the addition of prosodic knowledge extracted directly from the speech signal may prove helpful. The high correlation between pitch accents and semantic slots found in this study as well as the comparability of our results on ASR output motivates the use of prosody for slot filling. As future work, we aim to use pitch accents as features for SLU on recognized text.

6. Conclusion

This study aimed to find evidence that prosody, specifically the presence of pitch accents, can provide useful information for SLU tasks. In a first experiment, we found that a pitch accent detector trained on part of the Boston Radio News Corpus does not require pre-transcribed data: the transcriptions can be obtained from ASR and still yield comparable results, which means we can readily include pitch accent detection in a SLU system. The second part of this study examined the correlation of accents and semantic labels (*slots*) on the ATIS corpus. We show that most of the words in the corpus that are labelled with slots also carry a pitch accent. This finding agrees with the expectation of such words that are important and often new in the discourse to be perceptually more prominent. Furthermore, we determined that many words that are pitch accented but not associated with slots also convey substantial information about the speaker's objective, and can point towards their intention or the general domain. We conclude that prosodic information constitutes a useful resource that can be used to improve and enhance the automatic understanding of speech.

7. Acknowledgements

We would like to thank Antje Schweitzer for her support. This work was funded by the German Science Foundation (DFG), Sonderforschungsbereich 732 *Incremental Specification in Context*, Project A8, at the University of Stuttgart.

8. References

[1] J. Pierrehumbert and J. Hirschberg, "The meaning of intonational contours in the interpretation of discourse," in *Intentions in Communication*, J. M. P. Cohen and M. Pollack, Eds. MIT Press, Cambridge MA, 1990, pp. 271–311.

[2] E. Selkirk, "Sentence prosody: Intonation, stress and phrasing," in *The handbook of phonological theory*, J. A. Goldsmith, Ed. Oxford: Blackwell, 1995, pp. 550–569.

[3] ——, "The syntax-phonology interface," in *The handbook of phonological theory, Second Edition*, R. J. A. Goldsmith and A. C. L. Yu, Eds. Oxford: Blackwell, 2011, pp. 435–484.

[4] J. Hirschberg and J. B. Pierrehumbert, "The intonational structuring of discourse," in *24th Annual Meeting of the Association for Computational Linguistics, Columbia University, New York, New York, USA, July 10-13, 1986.*, 1986, pp. 136–144.

[5] A. Waibel, *Prosody and Speech Recognition*. Morgan Kaufmann, 1988.

[6] K. Vicsi and G. Szaszák, "Using prosody to improve automatic speech recognition," *Speech Communication*, vol. 52, no. 5, pp. 413–426, 2010.

[7] S. Ananthakrishnan and S. Narayanan, "Improved speech recognition using acoustic and lexical correlates of pitch accent in a n-best rescoring framework," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, Honolulu, Hawaii, 2007, pp. 873–876.

[8] J. H. Jeon, W. Wang, and Y. Liu, "N-best rescoring based on pitch-accent patterns," in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*. Portland, Oregon: Association for Computational Linguistics, 2011, pp. 732–741.

[9] K. Chen, M. Hasegawa-Johnson, A. Cohen, S. Borys, S.-S. Kim, J. Cole, and J.-Y. Choi, "Prosody dependent speech recognition on radio news corpus of American English," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 14, no. 1, pp. 232–245, 2006.

[10] A. Batliner, B. Möbius, G. Möhler, A. Schweitzer, and E. Nöth, "Prosodic models, automatic speech understanding, and speech synthesis: towards the common ground," in *Proceedings of the European Conference on Speech Communication and Technology (Aalborg, Denmark)*, vol. 4. ISCA, 2001, pp. 2285–2288.

[11] N. Veilleux and M. Ostendorf, "Prosody/parse scoring and its application in ATIS," in *In Proceedings of the ARPA Workshop on Human Language Technology*, 1993, pp. 335–340.

[12] E. Shriberg, A. Stolcke, D. Hakkani-Tür, and G. Tur, "Prosody-based automatic segmentation of speech into sentences and topics," *Speech Communication*, vol. 32, pp. 127–154, 2000.

[13] E. Shriberg and A. Stolcke, "Prosody modeling for automatic speech recognition and understanding," in *Mathematical Foundations of Speech and Language Processing*. Springer, 2004, pp. 105–114.

[14] J. Hirschberg, D. Litman, and M. Swerts, "Prosodic and other cues to speech recognition failures," *Speech Communication*, vol. 43, pp. 155–175, 2004.

[15] S. Goldwater, D. Jurafsky, and C. D. Manning, "Which words are hard to recognize? Prosodic, lexical and disfluendy factors that increase speech recognition error rates," *Speech Communication*, vol. 52, pp. 181–200, 2010.

[16] I. Rösiger and A. Riester, "Using prosodic annotations to improve coreference resolution of spoken text," in *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics (ACL-IJCNLP)*, Beijing, 2015, pp. 83–88.

[17] G. Tur and L. Deng, "Intent determination and spoken utterance classification," in *Spoken Language Understanding: Systems for Extracting Semantic Information from Speech*, G. Tur and R. D. Mori, Eds. John Wiley & Sons, 2011.

[18] C. T. Hemphill, J. J. Godfrey, and G. R. Doddington, "The ATIS Spoken Language Systems Pilot Corpus," in *Proceedings of the DARPA Speech and Natural Language Workshop*. Morgan Kaufmann, 1990, pp. 96–101.

[19] M. Ostendorf, P. Price, and S. Shattuck-Hufnagel, "The Boston University Radio News Corpus," Boston University, Technical Report ECS-95-001, 1995.

[20] F. Tamburini, "Prosodic prominence detection in speech," in *ISSPA2003*, 2003, pp. 385–338.

[21] K. Ross and M. Ostendorf, "Prediction of abstract prosodic labels for speech synthesis," in *Computer Speech & Language*, vol. 10, 1996, p. 155185.

[22] V. Kumar, R. Sridhar, S. Bangalore, and S. S. Narayanan, "Exploiting acoustic and syntactic features for automatic prosody labeling in a maximum entropy framework," in *IEEE Transactions on Audio, Speech & Language Processing*, vol. 16, no. 4, 2008, pp. 797–811.

[23] X. Sun, "Pitch accent prediction using ensemble machine learning," in *Proceedings of ICSLP-2002*, 2002, pp. 16–20.

[24] A. Schweitzer, "Production and perception of prosodic events-evidence from corpus-based experiments," Ph.D. dissertation, Universität Stuttgart, 2010.

[25] A. Rosenberg and J. Hirschberg, "Detecting pitch accents at the word, syllable and vowel level," in *HLT-NAACL*, 2009.

[26] ——, "Detecting pitch accent using pitch-corrected energy-based predictors," in *Proceedings of Interspeech*, 2007, pp. 2777–2780.

[27] G. Möhler and A. Conkie, "Parametric modeling of intonation using vector quantization," in *Proc. 3rd ESCA Workshop on Speech Synthesis, Jenolan Caves, Australia*, 1998, pp. 311–316.

[28] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, Nov. 2009.

[29] D. B. Paul and J. M. Baker, "The design for the Wall Street Journal-based CSR corpus," in *Proceedings of the workshop on Speech and Natural Language*. Association for Computational Linguistics, 1992.

[30] D. Povey, A. Ghoshal, G. Boulianne, L. Burget, O. Glembek, N. Goel, M. Hannemann, P. Motlicek, Y. Qian, P. Schwarz, J. Silovsky, G. Stemmer, and K. Vesely, "The Kaldi speech recognition toolkit," in *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding*. IEEE Signal Processing Society, Dec. 2011.

[31] J. Pitrelli, M. Beckman, and J. Hirschberg, "Evaluation of prosodic transcription labeling reliability in the ToBI framework," in *ICSLP*, 1994.

[32] N. T. Vu, "Sequential convolutional neural networks for slot filling in spoken language understanding," in *Proceedings of Interspeech*, 2016.

[33] N. T. Vu, P. Gupta, H. Adel, and H. Schütze, "Bi-directional recurrent neural network with ranking loss for spoken language understanding," in *Proceedings of the 41st IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Shanghai, China, 2016.

[34] G. Mesnil, Y. Dauphin, K. Yao, Y. Bengio, L. Deng, D. Hakkani-Tur, X. He, L. Heck, G. Tur, D. Yu, and G. Zweig, "Using recurrent neural networks for slot filling in spoken language understanding," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 23, pp. 530–539, 2015.