University of Stuttgart

Sabrina Stehwien, Ngoc Thang Vu

Institute for Natural Language Processing

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

Abstract

- prosody provides discourse information available from speech only
- pitch accents may help localize important words for SLU:

I'd LIKE to book a FLIGHT from SEATTLE to MUNICH

O O O O O O O B-fromloc.city_name O B-toloc.city_name

- investigate correlation between pitch accents and slots in ATIS
- simulate fully-automated SLU setup by using recognized text

Pitch Accent Detection from Audio Only

Model

- trained on subset of the Boston Radio New Corpus (1 hour 20 min)
- PaIntE and duration features in a random forest binary classifier

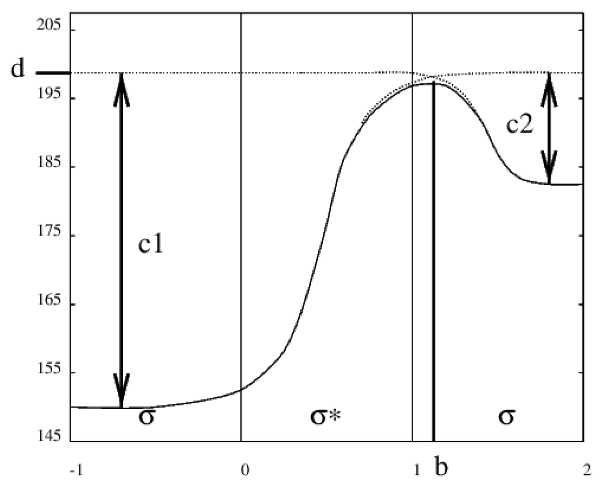


Figure 1: PaIntE [Möhler & Conkie, 1986] parameters describe the F0 contour

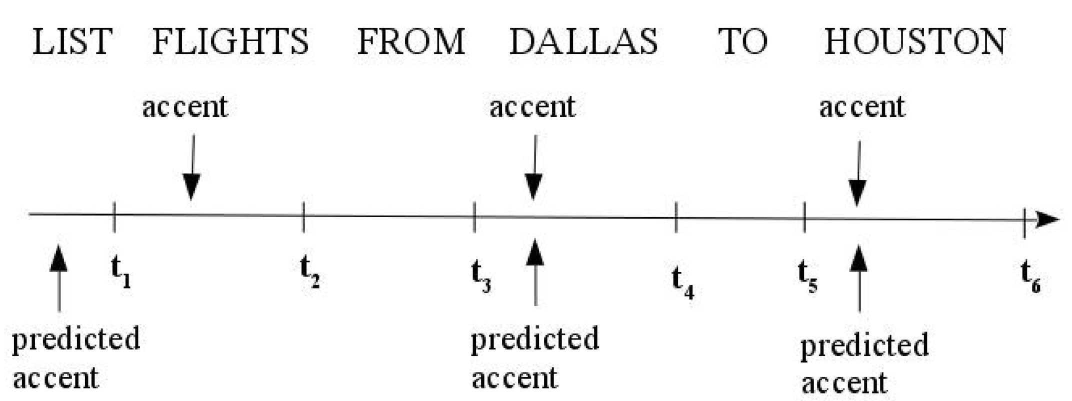
speaker	accuracy (%)
f1a	73.1
f2b	74.7
f3a	76.7
m1a	73.7
m2b	73.8

Table 1: speaker-independent pitch accent detection accuracy

Pitch accent detection on recognized text

- recognize same dataset with Kaldi (27% WER)
- apply speaker-independent pitch accent detection models as above
- evaluate against pitch accent labels in the Boston corpus:

Figure 2: we count a true positive if a predicted and a reference accent lie within the same time interval of a word in the reference text



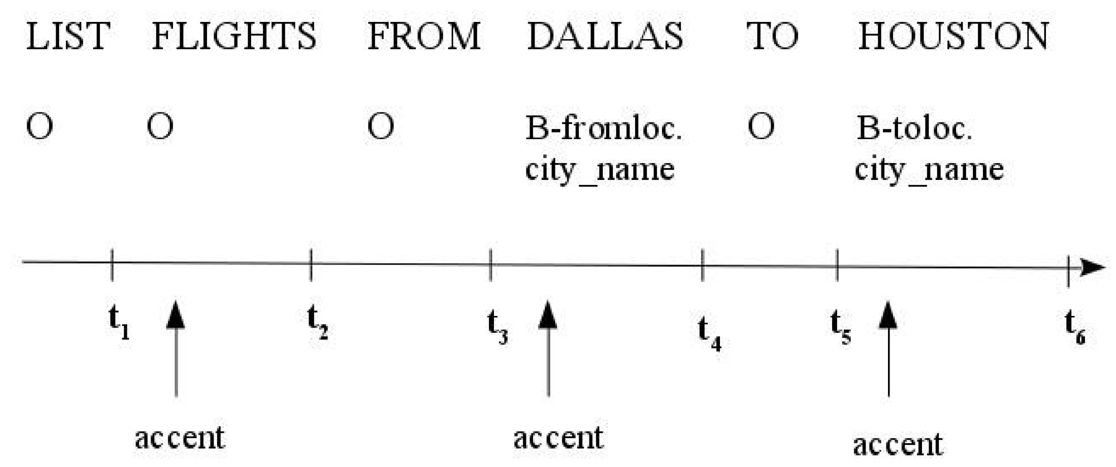
	reference	recognized
# predicted	7,460	8,520
# true positives	5,753	6,134
precision (%)	77.0	72.0
recall (%)	72.9	77.8
F1-Score (%)	74.9	74.8

Table 2: word-level accuracies of pitch accent detection using recognized and reference transcriptions of the Boston corpus; averages across speakers; corpus statistics: 220 files, 13836 words, 7888 accents in reference

Correlation of Pitch Accents and Semantic Slots in ATIS

- 607 utterances from ATIS3 test set recognized with 11.7% WER
- pitch accents predicted by model trained on all speakers in the Boston dataset

Figure 3: we consider how many times a pitch accent lies within the time interval of a word annotated with a slot label



Results

on original transcriptions:

# 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%

on recognized text:

# files	607
# words	6,212
# slots	2,452
# predicted accents	3,410
# pred. accents on slots	2,173
# pred. accents on non-slots	1,237
slots with pred. accent	88.7%

Tables 3 and 4: frequency of predicted pitch accents and coverage of slots in ATIS on the original transcriptions (left) and recognized text (right)

Pitch accented words not associated with slots:

- question words, imperatives *list, what, please, show, need* show speaker's intention
- function words are not frequently pitch accented (expected)
- domain indicators: *flight, flights* (accented only aroung 60% of the time, may be considered given)

word	accented (%)	frequency
WHICH	94.1	32
PLEASE	93.9	31
WHAT	90.8	79
LIST	87.3	137
NEED	78.0	39
ALL	77.1	37
SHOW	75.9	63
ME	74.5	76
I	73.9	78
FLIGHT	68.4	93
FLIGHTS	57.2	179
ON	51.9	83
AND	42.5	31
FROM	17.4	78
TO	17.3	87
THE	15.8	29

Table 5: most frequent accented non-slot words in ATIS

Agreement with Human Labelling

- estimate quality of pitch accent prediction on ATIS
- 50 ATIS utterances annotated with pitch accents by a human labeller
- agreement in over 70% of cases

# files	50
# words	514
# slots	201
# human-labelled accents	235
# words with predicted accents	234
agreement: # words	173

6

# **human-labelled** accents	235
# accents on slots	149
# accents on non-slots	86
# slots with no accent	52
slots with accent	74.1%

7

# **predicted** accents	234
# accents on slots	164
# accents on non-slots	70
# slots with no accent	37
slots with accent	81.6%

8

Tables 6-8: correlation between slots and human-labelled (left) and automatically predicted (right) accents in 50 ATIS files; above: agreement between the two

Conclusion

- the pitch accent detector trained on part of the Boston corpus does not require pre-transcribed data while yielding comparable results
  - → we can incorporate pitch accent detection in a SLU system
- most words in the ATIS corpus that are labelled with slots are pitch accented
  - → expectation of important words to be perceptually prominent
- many words that are pitch accented and are not labelled with slots also convey relevant information for SLU

Future Work

- different datasets necessary to test generalizability of results
- pitch accent features may help improve slot filling on ASR output, where performance drops [Mesnil, 2015]
- investigate whether there are correlations between different ToBI pitch accent types and/or different slot label types

Acknowledgements:
We would like to thank Antje Schweitzer for her support. This work was funded by the German Science Foundation (DFG), SFB 732, project A8, at the University of Stuttgart.

References:
- G. Mesnil, Y. Dauphin, K. Yao, Y. Bengio, L. Deng, D. Hakkani-Tur, X. He, L. Heck, G. Tur, D. Yu, G. Zweig, "Using recurrent neural networks for slot filling in spoken language understanding", in *IEEE Transactions on Audio, Speech and Language Processing,* pp. 530-539, 2015
- G. Möhler and A. Conkie, "Parametric modeling of intonation using vector quantization", in *Proceedings of the third ESCA Workshop on Speech Synthesis,* pp.311-316, 1998
- M. Ostendorf, P. Price, S. Shattuck-Hufnagel, "The Boston University Radio News Corpus", 1995
- J. Pierrehumber and J.Hirschberg, "The intonational structuring of discourse", in *24th Annual Meeting of the ACL*, pp. 136-144, 1986
- A. Schweitzer, "Production and perception of prosodic events – evidence from corpus-based experiments", Ph.D. dissertation, University of Stuttgart, 2010