IZVORNI ZNANSTVENI RAD
HTTPS://WWW.DOI.ORG/10.17234/9789533791562.04

Бранислав Геразов,[1] Веселинка Лаброска,[2] Ирена Савицка[3]

Универзитет „Св. Кирил и Методиј“ во Скопје, РС Македонија
[1]Факултет за електротехника и информациски технологии,
[2]Институт за македонски јазик „Крсте Мисирков“
[3]Институт за славистика, Полска академија на науките, Варшава, Р Полска
gerazov@feit.ukim.edu.mk, labroska_v@yahoo.com, irena.sawicka@ispan.waw.pl

Кодирање на фокусот на зборот со помош на интонацијата во македонскиот јазик

Со напредокот на говорната технологија, синтезата и препознавањето на фокусот на зборовите добија значење во подобрувањето на интерфејсот помеѓу луѓето и компјутерите, како и во системите за преведување од еден говор во друг. За да се олеснат овие алгоритми, важно е да се анализираат значењата што фокусот ги истакнува во говорот. Главниот канал што се користи за пренесување информации за фокусот на зборот е прозодијата. Различни јазици го прават тоа користејќи различни димензии на прозодија, главно преку интензитет, интонација и ритам. Во овој труд ги истражуваме разликите во динамиката на контурите F_0 што се јавуваат во македонскиот јазик кога еден од зборовите ќе се стави во фокус. Резултатите покажуваат дека интонацијата е важен показател за фокусирање во комуникацијата во македонскиот јазик.

Клучни зборови: прозодија, интонација, $Ф_0$, фокус, македонски јазик

1. Вовед

Прозодијата е важен канал во пренесувањето на лингвистичките информации (Cutler et al., 1997), на пр. структурата на реченицата, фокусот и контрастот, зборовниот акцент, но и паралингвистички информации (Schuller & Batliner, 2013), како полот, возраста, личноста и емоциите. Поради својата важност, прозодијата е клучна за развојот на говорните технологии, особено оние за синтеза на текст во говор, а помалку за автоматско препознавање на говор. Неодамна, прозодијата доби зголемена важност поради поместувањето на интересот на истражувањето кон областите на препознавање на

емоции во говорот (Vogt et al., 2008), синтезата на говор со емоции (Burkhardt & Campbell, 2014), како и синтезата на експресивен говор воопшто (Skerry-Ryan et al., 2018).

Фокусот на зборовите е важна лингвистичка функција која првенствено се реализира преку прозодијата. Поставувањето на фокусот во реченицата директно влијае на нејзиното значење и го модулира (Минова-Ѓуркова, 2000: 90; Савицка и сор., 2021: 117). Ова придонесува откривањето на фокусот на зборовите и самата синтеза да бидат многу важни за интерфејсите човек- компјутер (Clark et al., 2018), како што се интелигентните виртуелни лични асистенти: Алекса на Амазон, Сири на Епл (Apple), Асистентот на Гугл и Кортана на Мајкрософт. Овие системи можат да користат детекција на фокусот на зборовите за да заклучат кои информации ги нагласува говорникот или да користат синтеза со фокус за да го свртат вниманието на корисникот кон клучниот збор во синтетизираниот одговор. Друго важно апликативно сценарио се системите за преведување говор во говор (S2ST), кои денес стануваат реалност, на пр. преведувач на Скајп (Skype). Системите S2ST идеално треба да ја одредат важноста на секој збор во изворниот јазик (L1) и да ја пренесат оваа информација на целниот јазик (L2) (Anumanchipalli et al., 2012; Do et al., 2015).

Иако денес истражувањето на прозодијата се прошири за да ги покрие и визуелните знаци, како што се изразите на лицето и говорот на телото (Krahmer & Swerts, 2009), трите главни димензии на говорната прозодија остануваат да бидат интонацијата, интензитетот и ритамот. Се покажа дека кодирањето на фокусот на зборот е означено со секој од овие елементи: интонација (Ladd & Morton, 1997), интензитет (енергија) (Cernak & Honnet, 2015; Heldner et al., 1999; Stojkovic et al., 2015), ритам (Melov et al., 2015), како и со сите нив заедно (Gerazov et al., 2016; Terken & Hermes, 2000).

Средствата со кои прозодијата се користи за да се пренесе фокусот на зборот се разликуваат од јазик до јазик. Во некои јазици фокусот на зборот најчесто се кодира преку интонацијата, додека во други повеќе се потпира на ритамот или интензитетот. Освен тоа, кое од овие средства ќе доминира може да зависи и од говорникот. Затоа, важно е да се разјасни како фокусот е кодиран во конкретниот јазик што е предмет на наша анализа. Во овој труд ја истражуваме

варијабилноста на контурата на F_0 условена од присуството на фокус на зборот во стандардниот македонски јазик.

2. Фокус на зборот

Во неутрални искази, последната стапка обично го носи прозодиското обележување, т.е. граничниот тон на интонациската контура. Исказите со фокус на даден збор од фразата се разликуваат по тоа што значителна важност му се дава на различен збор во исказот. Дури и кога зборот во фокус е последниот збор од фразата, постои контраст во динамиката на прозодијата што го разликува исказот од неутралниот.

Фокусот на зборот се користи за менување на функционалната перспектива на пораката вградена во исказите; воведува модална содржина, на пр. неверување или емоционално обележување. Фокусот на зборот е задолжителен во фрази со специфична морфолошка структура, како што се прашањата со морфолошки ознаки каде што обично се нагласува прашалниот збор. Конечно, фокусот на зборот може да го промени семантичкото значење на исказот.

На пример, реченицата „Тој зборуваше.” во својата неутрална форма, носи реченичен акцент на глаголот, што го прави малку поистакнат. Ова му пренесува на слушателот информација дека некој зборувал, притоа не пренесувајќи ништо необично за ситуацијата. Спротивно на тоа, реченицата „Тој **зборуваше**!” ќе изрази изненадување дека некој навистина зборувал. Алтернативно, „**Тој** зборуваше!” ќе го изрази изненадувањето на говорникот за лицето што зборувало, т.е. конечно, фокусот на зборот може да се користи за да се пренесе контраст помеѓу два збора во една реченица, на пр. „Тој сака да **слика**, а не да **црта**!”

3. Експерименти

За да ја анализираме реализацијата на фокусот на зборовите на македонски јазик, снимивме мал сет на искази во кои беа модулирани три носечки фрази со помош на фокусот на зборот. Распределбата на податоците е прикажана во Табела 1. Првите две носечки фрази се исказни, а последната е прашање. Во Табелата се наведени маке-

донските верзии на трите фрази, нивната фонетска транскрипција со помош на Меѓународната фонетска азбука (IPA) и нивниот англиски превод во кој го задржуваме оригиналниот редослед на зборовите. Подвлечените зборови беа ставени во фокус во различните реализации на исказите. За секоја реченица е снимен и неутрален исказ.

Снимките се направени со говорник на македонски од машки пол што живее во Скопје и кој секојдневно професионално користи стандарден македонски јазик. Авторите внимаваа во сите снимки на говорникот да биде истакнат бараниот фокус на зборот или неговото отсуство. Снимките се направени во професионалното говорно студио на Факултетот за електротехника и информациски технологии во Скопје, со помош на кондензаторски микрофон и надворешен аудио интерфејс. Снимките беа семплирани со фреквенција од 44,1 kHz и квантизирани на 16 бита.

Интонацијата на исказите беше извлечена со помош на алгоритамот за екстракција на основната фреквенција имплементиран во Пакетот со алатки за препознавање говор Калди (Ghahremani et al., 2014). Алгоритамот дава континуирана проценка на F_0 преку употреба на интерполација и измазнување. Исто така, дава проценка на веројатноста за звучност (POV), која ја користевме за соодветно да ги означиме звучните делови од извлечените контури во нашите искази.

За да ги споредиме различните реализации на иста носечка фраза, ги усогласивме нивните акустични реализации врз основа на нивната спектрална содржина. Ова беше реализирано со помош на алгоритмот за динамичко искривување на времето (DTW) (Rabiner & Juang, 1993) имплементиран во Python модулот dtw[1]. Временското искривување беше засновано на мел- фреквентните цепстрални коефициенти (MFCCs) (Davis & Mermelstein, 1980) извлечени со помош на библиотеката librosa (McFee et al., 2015).

1 https://github.com/pierre-rouanet/dtw

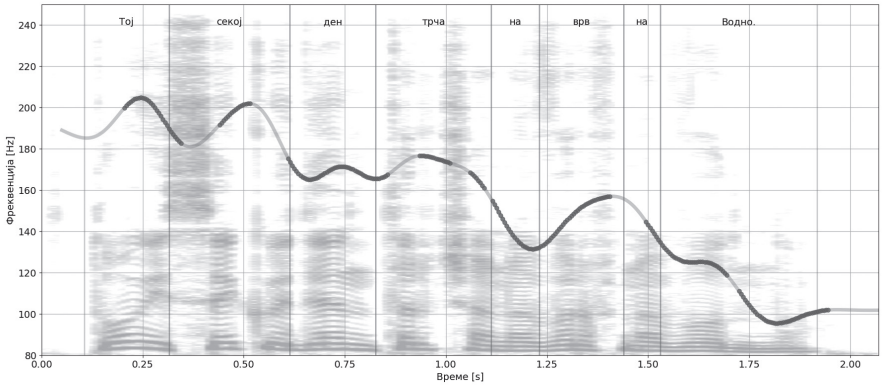
Табела 1. Распределба на податоците во анализата

Носечка фраза	број на снимања		
	неутрално	збор во фокус	вкупно
Тој секој ден трча на врв на Водно. /toj sekoj den tr̩tʃa na vr̩v na vodno/ He every day runs to the peak of Vodno.	1	5	6
Студов секогаш доаѓа од планиниве. /studov sekogaʃ doaja od planinive/ The cold always comes from these mountains.	1	4	5
Дали Никола е подобар од Горан? /dali nikola e podobar od goran/ Is Nikola better than Goran?	1	3	4
Вкупно	3	12	**15**

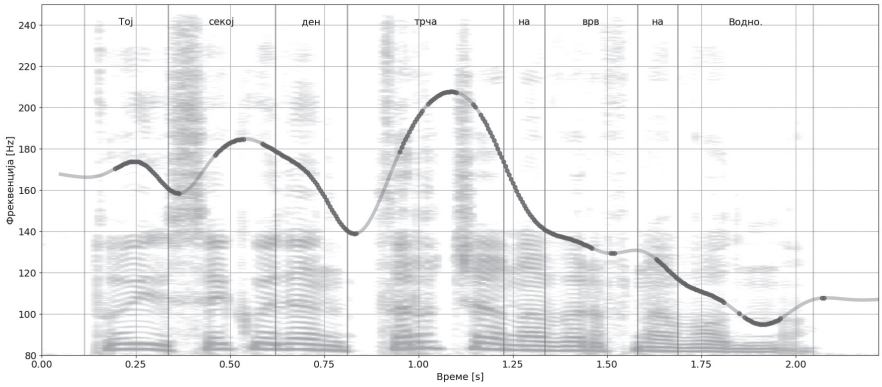
4. Резултати

Добиените контури на F_0 од два примери на реализација на првата носечка фраза се прикажани во врз нивните спектрограми на сл. 1 и 2. Овде, сл. 1 покажува неутрална реализација, додека сл. 2 прикажува реализација со фокус на зборот „трча". Можеме да видиме дека во неутралната реализација има мал пораст на интонациската контура за акцентираниот слог на секој од содржинските зборови „тој", „секој", „ден", „трча", „врв" и „водно". Тие кореспондираат со движењата на тонот L*+H според ToBI системот за означување (Silverman et al., 1992), а нивната крива е во форма на ѕвонче, т.е. во низа ниско-високо-ниско и се нарекува уште *реверс патерн* (*reverse pattern;* Lehiste & Ivić, 1980). Сепак, овие мали пертурбации во F_0 не се тонски акценти, туку поверојатно корелации на лексичкиот акцент. Проверката на ова е надвор од опсегот на оваа анализа.

Од друга страна, во реализацијата со фокус на зборот, во F_0 е јасно означена контурата за истакнатиот збор. Повторно гледаме тонски акцент во L*+H или *реверс патерн*, но овој пат многу поизразен во однос на динамиката на F_0 на остатокот од исказот.



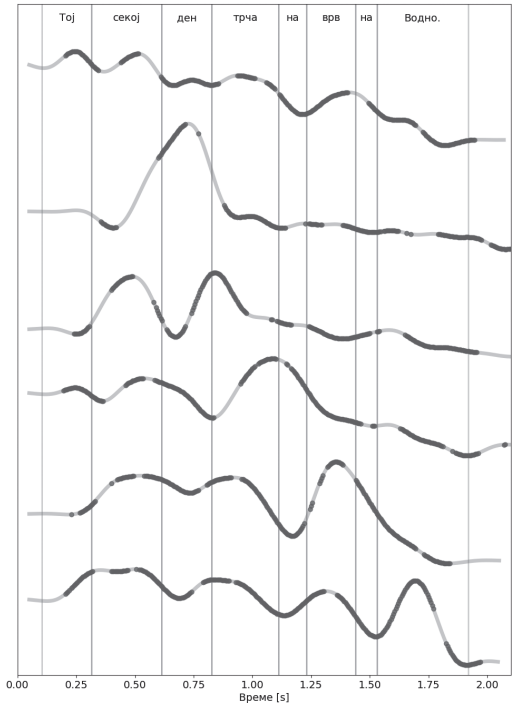
Слика 1. Контурата F_0 при неутрална реализација на фразата „Тој секој ден трча на врв на Водно.”



Слика 2. Контурата F_0 на фразата „Тој секој ден трча на врв на Водно.” со фокус на зборот „трча”

Временски усогласената споредба на сите остварувања на двата исказа (двете фрази) е прикажана на сл. 3 и 4. Можеме да видиме дека набљудуваниот *реверс патерн* (обратна шема) што се користи за означување на зборот во фокус на сл. 2, се пренесува на сите фокусирани зборови во двете фрази. Интересно е да се забележи дека во реализацијата со фокус на зборот „секој” на сл. 3, обратната шема (реверс патерн) се протега на следниот збор, што нè наведува да веруваме дека тонскиот акцент L*+H е оној што ја дава информацијата

на зборот во фокус, наместо самиот реверс патерн. Ова може делумно да се забележи и во следната реализација со фокус на зборот „ден". Во оваа реализација можеме да забележиме и тонски акцент на „секој" кој е помалку изразен од оној на „ден".

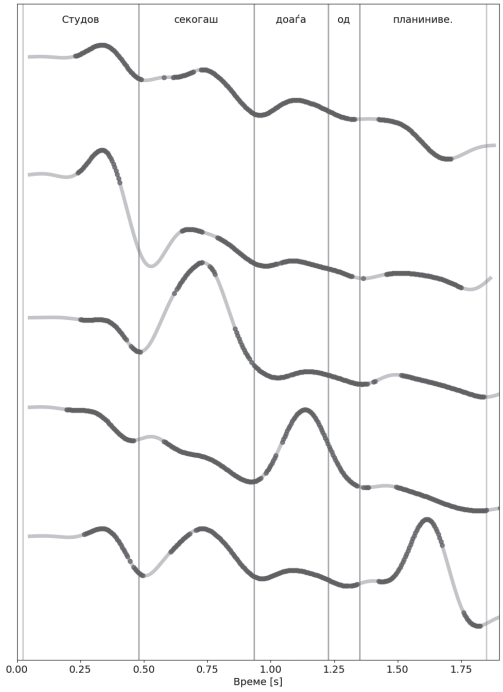


Слика 3. Споредба на контурите F_0 за различните реализации на фразата „Тој секој ден трча на врв на Водно.", почнувајќи од неутралната реализација (најгоре), и продолжувајќи со зборовите во фокус како што следува: „секој", „ден", „трча", „врв" и „Водно"

Конечно, можеме да видиме дека сите фрази имаат контури со пад на крајот, т.е. се работи за L-L% фраза и граничен тон. Исто така, последниот збор од исказите не е посебно означен во отсуство на збор во фокус и покажува постепено намалување на F_0 во текот на целото негово времетраење.

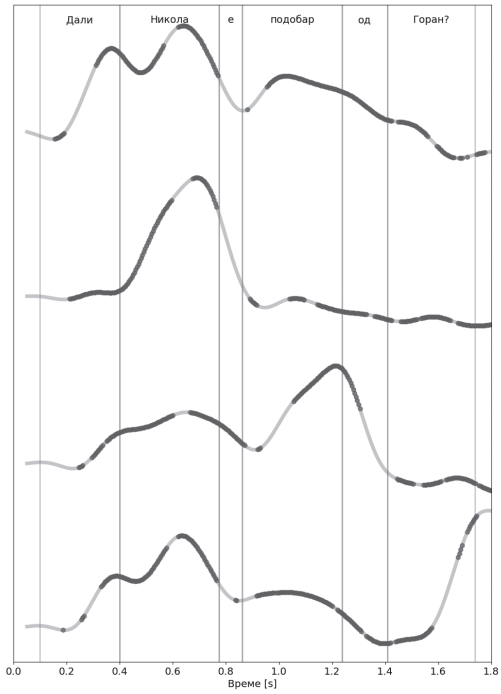
Од сл. 4 може да се види дека можеме да ги донесеме истите заклучоци за F_0 како кај сл. 3, со таа разлика што овде сите зборови во фокус се поистакнати отколку кај погорната реченица. Ова можеби

се должи на фактот дека реченицата е пократка па говорникот може со поизразена артикулација да го означи фокусот на сите позиции.



Слика 4. Споредба на контурите F_0 за различните реализации на фразата „Студов секогаш доаѓа од планиниве.” почнувајќи од неутралната реализација (најгоре), и продолжувајќи со зборовите во фокус како што следува: „студов”, „секогаш”, „доаѓа” и „планиниве”

Различните остварувања на носечката прашална фраза се прикажани на сл. 5. Во неутралната изведба на прашалната фраза, акцентот паѓа на зборот што следи по прашалниот збор „дали”, но има и изразена динамика на F_0 на самиот прашален збор. Тука повторно ги имаме L*+H тонските акценти и за нив поврзаните реверс патерн (обратни шеми), како што беше случајот и со фокусот на зборовите во исказните реченици.



Слика 5. Споредба на контурите F_0 за различните реализации на фразата „Дали Никола е подобар од Горан?" почнувајќи од неутралната реализација (најгоре), и продолжувајќи со зборовите во фокус како што следува: „Никола", „подобар" и „Горан"

За разлика од неутралната реализација, кога фокусот на зборот се става на „Никола", тонскиот акцент на „дали" исчезнува, а „Никола" станува единствениот доминантен тонски акцент во исказот. Конечно, можеме да видиме дека првите три реализации имаат L-L% фразен и граничен тон, што не е невообичаено за морфолошки означени прашања.

Меѓутоа, кога фокусот на последниот збор „Горан" бара акцент на тонот L*+H, нема паѓање на F_0 назад на L%, туку наместо тоа, се појавува граничен тон H%. Ова го потврдува нашиот претходен заклучок дека L*+H е вистинскиот маркер на фокусот на зборовите и дека појавата на реверс патерн (обратна шема) е само страничен ефект на интонациската контура која се враќа на нејзината поранешна ниска вредност на F_0.

5. Заклучок

Од презентираната кратка анализа можеме да заклучиме дека постои силно означување на фокусот на зборовите со помош на F_0 во македонскиот јазик. Тоа се реализира со нагласена динамика на F_0 на зборот во фокус, надополнета со потиснување на динамиката на висината за остатокот од исказот. Така, говорникот јасно ја обележува важноста на зборот преку неговиот обемен ларингален напор и недостатокот од истиот во текот на остатокот од изговорот. Карактеристичниот тонски акцент што се користи за кодирање на зборот во фокус на македонски јазик е тонот L*+H според ToBI, или поконкретно острото зголемување на F_0 на акцентираниот слог на зборот во фокус. Оваа информација е важна и за синтеза и за препознавање на зборовите во фокус во системите за македонски јазик како и за системите за превод од говор на говор во кој еден од јазиците ќе биде македонскиот стандарден јазик.

6. Благодарност

Ова истражување е финансирано со грант од Националниот научен центар во Полска, Проект бр. 2017/25/B/HS2/00760.

7. Литература

кирилица:

Минова-Ѓуркова, Л. (2000). *Синтакса на македонскиот стандарден јазик*. Магор.

Савицка, И., Геразов, Б., Лаброска, В., Цихнерска, А. и Травињска, А. (2021). *Фонетика и фонологија на македонскиот стандарден јазик. Супрасегментална фонетика и фонологија*. МАНУ.
http://ical.manu.edu.mk/books/Suprasegmentalna_fonologija.pdf

латиница:

Anumanchipalli, G. K., Oliveira, L. C., & Black, A. W. (2012). Intent transfer in speech-to-speech machine translation. *Proceedings of the fourth IEEE Workshop on Spoken Language Technology* (153–158).

Burkhardt, F., & Campbell, N. (2014). Emotional speech synthesis. In R. Cal-

vo, S. D'Mello, J. Gratch, & A. Kappas (Eds.), *The Oxford Handbook of Affective Computing*: 286. Oxford University Press.

Cernak, M., & Honnet, P. E. (2015). An empirical model of emphatic word detection. In *Proceedings of Interspeech*, Dresden, Germany, September.

Clark, L., Doyle, P., Garaialde, D., Gilmartin, E., Schlögl, S., Edlund, J., Aylett, M., Cabral, J., Munteanu, C., & Cowan, B. (2018). The state of speech in HCI: Trends, themes and challenges. *arXiv preprint arXiv:1810.06828*.

Cutler, A., Dahan, D., & Van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, *40*(2), 141–201.

Davis, S. B., & Mermelstein, P. (1980). Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. *IEEE Trans. Acoust. Speech Signal Process*, *28*(4), 357–366.

Do, Q. T., Takamichi, S., Sakti, S., Neubig, G., Toda, T., & Nakamura, S. (2015). Preserving word-level emphasis in speech-to-speech translation using linear regression HSMMs. *Proceedings of Interspeech*, Dresden, Germany.

Gerazov, B., Gjoreski, A., Melov, A., Honnet, P. E., Ivanovski, Z., & Garner, P. N. (2016). Unified prosody model based on atom decomposition for emphasis detection. *ETAI,* Struga, Macedonia, Sep 22–24, 2016.

Ghahremani, P., BabaAli, B., Povey, D., Riedhammer, K., Trmal, J., & Khudanpur, S. (2014). A pitch extraction algorithm tuned for automatic speech recognition. *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE: 2513–2517.

Heldner, M., Strangert, E., & Deschamps, T. (1999). A focus detector using overall intensity and high frequency emphasis. *Proceedings of the International Congress of Phonetic Sciences (ICPhS)*, Vol. 2, 1491–1493.

Krahmer, E., & Swerts, M. (2009). Audiovisual prosody–introduction to the special issue. *Language and Speech*, *52*(2–3), 129–133.

Ladd, D. R., & Morton, R. (1997). The perception of intonational emphasis: Continuous or categorical? *Journal of Phonetics*, *25*(3), 313–342.

Lehiste, I., & Ivić, P. (1980). The intonation of yes-or-no questions – A new Balkanism? *Balkanistica*, *6*, 45–53.

McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., & Nieto, O. (2015). librosa: Audio and music signal analysis in python.

Proceedings of the 14th Python in Science Conference, Vol. 8.

Melov, A., Gerazov, B., & Ivanovski, Z. (2015). Emphatic word detection based on syllable durations. *XII International Conference ETAI*, Ohrid, Macedonia.

Rabiner, L., & Juang, B. (1993). *Fundamentals of speech recognition*. Prentice Hall.

Schuller, B., & Batliner, A. (2013). *Computational paralinguistics: Emotion, affect and personality in speech and language processing*. John Wiley & Sons.

Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., Pierrehumbert, J., & Hirschberg, J. (1992). ToBI: A standard for labeling English prosody. *The Second International Conference on Spoken Language Processing*, CSLP 1992, Banff, Alberta, Canada, October 13–16.

Skerry-Ryan, R. J., Battenberg, E., Xiao, Y., Wang, Y., Stanton, D., Shor, J., Weiss, R. J., Clark, R., & Saurous, R. A. (2018). Towards end-to-end prosody transfer for expressive speech synthesis with Tacotron. *arXiv preprint arXiv:1803.09047*.

Stojkovic, A., Gerazov, B., & Ivanovski, Z. (2015). Emphatic word detection based on relative phoneme energies within syllables. *XII International Conference ETAI*, Ohrid, Macedonia, Sep 24–26, 2015.

Terken, J., & Hermes, D. (2000). The perception of prosodic prominence. In M. Horne (Ed.), *Prosody: Theory and experiment, studies presented to Gösta Bruce* (89–127). Kluwer.

Vogt, T., André, E., & Wagner, J. (2008). Automatic recognition of emotions from speech: A review of the literature and recommendations for practical realisation. In C. Peter, & R. Beale (Eds.), *Affect and emotion in human-computer interaction. Lecture notes in computer science*, Vol. 4868. Springer.

Coding of word focus using intonation in the Macedonian language

With the advancement of speech technology, the synthesis and recognition of word focus has gained importance in advancing human-computer interfaces, as well as speech-to-speech translation systems. To facilitate these algorithms, it is important to analyse the means by which focus is communicated in speech. The main channel used to transfer the information of word focus is prosody. Different languages do this using different dimensions of prosody, mainly through intensity, intonation, and rhythm. We explore the differences in the dynamics of the F0 contour that occur in Macedonian, when a word is placed in focus. Results show that intonation is an important marker in communicating focus in Macedonian.

Keywords: prosody, intonation, F_0, focus, Macedonian