

Can speech pitch perception be measured language-independently?

Willemijn Heeren^a, Andrei A. Avram^b, Anna Cardinaletti^c,
Martine Coene^d and Francesca Volpato^c

^aLeiden University / ^bBucharest University / ^cCa' Foscari University of Venice / ^dUniversity of Antwerp

Recently, a test battery was developed with the goal of assessing perception of F_0 in linguistic, but language-independent, contexts by listeners from different language backgrounds. Test validation using Dutch, Italian and Romanian listeners generally showed comparable performance. In this study we present cross-language results on how F_0 contributes to prominence perception in Dutch (a Germanic language) and Italian and Romanian (Romance languages), as well as on how F_0 trades off with duration, a second important cue to prominence. We discuss implications of these results for the language-independent set-up of the test battery.

Keywords: pitch perception, cue trading, prominence marking

1. Introduction

Recently, a test battery was developed to assess pitch perception in speech (Heeren et al. in press). The tests are intended to support the increased attention paid to prosody in clinical audiology practice, and to help assess new technologies for improved pitch perception in listeners with profound hearing losses. Earlier tests assess pitch perception by presenting tone complexes, which may not be sufficiently representative of speech processing, or natural speech recordings, which may in fact reflect reliance on secondary cues. The new test battery measures just noticeable differences (JNDs) for changes in fundamental frequency (F_0) embedded in linguistic contexts, without confounding secondary acoustic cues. The linguistic contexts were designed for listeners from different language backgrounds (Romance and Germanic languages), by using a pseudo-language to build stimulus materials and by basing tasks on linguistic phenomena that rely on F_0 as a cue

in several western European languages viz. question versus statement intonation, and prominence marking.

Test validation had shown highly comparable performance for listeners from different language backgrounds (Dutch, Italian, Romanian) on the question/statement task, but when asked to identify the most prominent syllable using F_0 alone, Italian listeners in some respects differed from Dutch and Romanian listeners. Languages differ in their use of and trade-off between acoustic cues, so small between-language differences can be expected. The goal of this investigation was to study how the language-independent set-up of the test battery (pseudo-language, stylized pitch) compared to native language processing.

1.1 Background

A prominence-lending pitch accent can contribute to meaning both at the word level, i.e. lexical stress, and at the phrase level, i.e. focus. For Dutch, words with a prominence-lending pitch accent are longer than unaccented ones, where lengthening is found for each syllable in the accented word (Eefting 1991). Relative syllable duration has been reported as the most reliable correlate of lexical stress when a word is placed in sentence context (Sluijter & van Heuven 1996). In perception, F_0 is taken to be the primary cue to stressed syllables in Dutch (van Katwijk 1974), as was also reported for English (Fry 1958), but also duration and spectral balance are perceptual cues for Dutch listeners (Sluijter et al. 1996). A pitch accent is furthermore perceived as signaling a focused constituent (Nooteboom & Kruyt 1987).

Italian has many varieties that exhibit differences at both the segmental and supra-segmental levels (Rossi 1998), but we focused on Northern and Standard Italian (cf. Bertinetto & Loporcaro 2005). (Non-final) stressed syllables are produced longer than unstressed ones (Bertinetto 1981; Nespor & Vogel 1986; Avesani et al. 2007). Intensity and F_0 were not found to consistently change with stress position, when measured in minimal stress triplets such as *'capito — ca'pito — capi'to* (Bertinetto 1981). This also held for vowel quality (Bertinetto 1980). Nouns with narrow focus elicited from four Northern Italian speakers were in all cases associated with a pitched movement, and in most cases a rise-fall on the lexically stressed syllable (Farnetani & Zmarich 1997). Contrastive focus was also realized through a rising-falling pitch movement on the stressed syllable, and it furthermore showed increased duration and intensity (Bertinetto & Loporcaro 2005).

In perception, the duration cue has been reported as the main one for lexical stress perception (Bertinetto 1980; Alfano 2006). As to the contribution of pitch accents, reports are contradictory. Bertinetto (1980) used the synthetic minimal stress pair *'papa — pa'pa* (pope — father), and varied duration, intensity and F_0 . Listeners associated lengthening with the presence of stress, and the author also

concluded that F_0 was a 'relatively weak' cue. Alfano (2006) confirmed that duration was the main cue, but especially when combined with F_0 .

In Romanian, stressed vowels are longer than their unstressed counterparts (Giurgiu 2008; Manolescu et al. 2009). Minimal pairs of words, such as '*casa* — *ca'sa*' (house — to quash), furthermore showed increased mean F_0 and intensity for stressed syllables, but this data set was small and results were not statistically analyzed (Giurgiu 2008). Manolescu et al. (2009) studied whether vowel duration and pitch accents were used to signal contrastive focus and found that duration did not seem to be a reliable cue to contrastive focus, but that F_0 was. As for perception, Avram (1966) concluded that vowel duration was not relevant in stress perception, and later reported that F_0 contributed to stress perception (Avram 1970).

1.2 The present research

To answer the question if pitch perception can be measured language-independently in the test battery, we investigated two issues.

The first question was how F_0 and duration trade off as cues to prominence perception in each of the three languages (Dutch, Italian, Romanian), and especially how listeners respond to natural F_0 changes when duration is neutralized. This is a test of the task's underlying assumption that F_0 cues prominence. It is expected that duration may be used less by Romanian than by Dutch and Italian listeners, and that pitch accents are associated with prominence by Dutch and Romanian listeners, whereas this association is expected to be less strong for Italian listeners. In the case of conflicting cues, F_0 cues are expected to overrule duration cues for Dutch listeners, but the expectation is the other way around for Italian listeners.

The second question was if the F_0 manipulations used in the test battery are perceived as intended when applied to the listeners' native language rather than a pseudo-language. The task was not to indicate the most prominent syllable, as in the test battery, but the more natural task of which word meaning listeners perceived. The expectation is that listeners perceive prominence as predicted when the stylized pitch accent is used in words of their own language.

2. Method

A two-category identification task was used in which listeners made a semantic judgment by indicating which member of a native language minimal stress pair they heard. Both duration and F_0 were manipulated so that they (1) co-indicated the stressed syllable (lengthening and pitch accent on the same syllable), (2) provided conflicting information as to which syllable was stressed (lengthening

on one and F_0 marking on the other), or (3) gave single-cue information (F_0 or duration).

2.1 Recordings

Per language, a two-syllable minimal stress pair was selected: *'papa — pa'pà* (pope — father) for Italian, *'copii — co'pii* (copies — children) for Romanian, and *'canon — ka'non* (canon — cannon) for Dutch. Minimal stress pairs were recorded in both focused and non-focused settings according to the Sluijter & Van Heuven (1996) sentence paradigm, translated into each of the three languages:

Italian

- | | |
|---|--|
| a. Per favore dica <u>papa</u> invece di vescovo. | please say <u>pope</u> instead of bishop |
| b. Per favore <u>dica</u> papa invece di scriverlo. | please <u>say</u> pope ... writing it down |
| c. Per favore dica <u>papà</u> invece di mamma. | please say <u>father</u> ... mother |
| d. Per favore <u>dica</u> papà invece di scriverlo. | please <u>say</u> father ... writing it down |

Romanian

- | | |
|--|--|
| a. Te rog zi <u>copii</u> în loc de originale. | please say <u>copies</u> instead of originals |
| b. Te rog <u>zi</u> copii în loc să scrii. | please <u>say</u> copies ... writing it down |
| c. Te rog zi <u>copii</u> în loc de adulți. | please say <u>children</u> ... adults |
| d. Te rog <u>zi</u> copii în loc să scrii. | please <u>say</u> children ... writing it down |

Dutch

- | | |
|---|--|
| a. Wil je <u>canon</u> zeggen in plaats van liedje. | please say <u>canon</u> instead of song |
| b. Wil je <u>canon</u> <u>zeggen</u> in plaats van schrijven. | please <u>say</u> pope ... writing it down |
| c. Wil je <u>kanon</u> zeggen in plaats van geweer. | please say <u>cannon</u> ... rifle |
| d. Wil je <u>kanon</u> <u>zeggen</u> in plaats van schrijven. | please <u>say</u> cannon ... writing it down |

Three female speakers, one from each language background, were recorded individually in a sound-treated booth at the phonetics laboratory of Leiden University. Recordings were made directly onto a computer (44.1 kHz, 16 bits/sample), using a Sennheiser MKH 416T directional condenser microphone.

F_0 and duration measurements of the recordings revealed that the minimal pairs adhered to acoustic patterns reported in the literature (see Table 1). The duration ratios of the 1st- and 2nd-syllable vowels show that stress position affects duration. In Dutch and Italian productions, the stressed syllable has the longer vowel duration, in line with Bertinetto (1981) and Sluijter & Van Heuven (1996). Moreover, Dutch words in focus showed a time expansion of the entire word (e.g. Eefting 1991). For the Romanian words, vowel durations are comparable between the first and second syllable within a word when stress is on the first syllable, whereas the stressed vowel is longer than the unstressed one when stress falls on

the second syllable. Following the literature (Giurgiu 2008), stressed vowels were longer than their unstressed counterparts in the same position.

Non-focused productions showed smaller F_0 ranges and smaller differences between the mean F_0 s over the two syllables of a word than focused productions in all three languages. There were differences between pitch accent realizations associated with focus in the three languages; in Italian, the pitch range on the accented syllable was larger, whereas Romanian and Dutch productions showed a higher average F_0 on the stressed syllable.

Table 1. F_0 measurements in Hertz and vowel durations in milliseconds of the original recordings (S1F: 1st-syllable focused, S1NF: 1st-syllable non-focused, S2F: 2nd-syllable focused, S2NF: 2nd-syllable non-focused).

Language	Condition	Word form	Mean F_0		F_0 range		Duration		Duration Ratio
			s1	s2	s1	s2	V1	V2	V1/V2
Dutch	S1F	<u>ca</u> non	302	214	45	100	187	101	1.85
	S1NF	<u>ca</u> non	164	149	27	16	141	89	1.58
	S2F	ka <u>no</u> n	179	240	23	123	59	88	0.67
	S2NF	ka <u>no</u> n	161	159	4	7	48	80	0.82
Italian	S1F	<u>pa</u> pa	207	169	60	13	186	125	1.49
	S1NF	<u>pa</u> pa	168	169	16	8	177	102	1.74
	S2F	pa <u>pa</u>	211	185	35	50	72	118	0.61
	S2NF	pa <u>pa</u>	172	169	14	27	87	144	0.61
Romanian	S1F	<u>co</u> pii	265	223	11	46	121	109	1.11
	S1NF	<u>co</u> pii	200	193	34	3	128	132	0.97
	S2F	co <u>pi</u> i	233	249	29	2	95	157	0.61
	S2NF	co <u>pi</u> i	214	206	25	29	101	175	0.57

2.2 Stimulus materials

Differences in vowel quality between lexically stressed and unstressed versions of the same syllable were minimized to restrict their influence as possible cues, and syllable loudness was normalized within each word.

Stimuli consisted of a carrier sentence followed by a target word. Per language background, five types of stimuli were prepared.

1. The baseline stimuli were the original recordings after application of vowel quality and loudness normalization. Informal pilot tests with native listeners confirmed that these were perceived as intended.

2. Two types of stimuli with conflicting information were generated: (1) stimuli with original duration information, but pitch information borrowed from the other member of the minimal stress pair, and (2) stimuli with original pitch information, but duration borrowed from the other member of the minimal stress pair.

For the first type, the pitch contour was stylized per word form, and anchors were created at the onsets and offsets of syllables and at within-syllable locations where the pitch course changed direction. Pitch contours of the ‘donor’ were transferred (through linear interpolation) to corresponding anchors of the ‘recipient’. When the donor had 1st-syllable stress, the recipient always had 2nd-syllable stress, and vice versa. This resulted in 8 stimuli per language: 4 word forms \times 2 substitutions per word form.

For the second type of conflicting information stimuli, only duration profiles were exchanged between word forms. The duration ratio between the ‘donor’ and ‘recipient’ segments was computed, and applied as a factor for linear compression or expansion to the ‘recipient’. When the donor had 1st-syllable stress, the recipient always had the durational make-up suggesting 2nd-syllable stress, and vice versa. After duration manipulation, the original pitch contours were replaced onto the audio files to restore slight changes that had occurred as a result of duration changes through PSOLA re-synthesis (Moulines & Charpentier 1990). This resulted in 8 stimuli per language: 4 word forms \times 2 substitutions per word form.

3. The duration-only stimuli were made by maintaining the original recording’s duration profile, whereas the pitch contour was flattened to the mean value over the word. This resulted in 4 duration-only stimuli per language.
4. For the pitch-only stimuli, durations were normalized and pitch was maintained as in the original recordings (controlling for syllable length). Duration normalization meant durations of 100 ms for consonants and 150 ms for vowels in the CV syllables. The standardized durations for the Dutch final CVC syllable were derived from (1) a total syllable length of 250 ms, and (2) the mean segment durations across Dutch word forms and conditions. This resulted in a 60 ms consonant, followed by a 120 ms vowel and a 70 ms consonant. These manipulations gave 4 pitch only stimuli per language.
5. The stylized pitch stimuli had the normalized duration that was also used for the pitch-only stimuli, with pitch contours (H*L) as in the test battery: a peak at 50 ms into the vowel resulting from a rise starting 100 ms before, and followed by a 100 ms fall back to the baseline. Pitch maxima were set at either 2, 4 or 6 semitones from the baseline. This resulted in 24 stimuli per language: 4 word forms \times 2 stress locations \times 3 pitch excursion sizes.

All stimulus words were preceded by a neutral carrier sentence in their respective language (Italian: *Per favore dica...*; Romanian: *Te rog zi...*, both meaning 'Please say...'; Dutch: *Zeg nog eens...*, meaning 'Say once more...'). Each stimulus started with 150 ms of initial silence, which included the initial stop closures in the case of Italian and Romanian. If necessary, the level of the carrier sentence's pitch course was adjusted to match that of the target word; it was heightened/lowered to obtain a smooth continuation of F_0 where carrier and target were concatenated, while taking into account the declination of the sentence as a whole.

2.3 Procedure

A two-category classification task was used with the two members of the minimal stress pair as response categories. Listeners responded which word, i.e. which meaning, they perceived. A total of 82 self-reported normal-hearing adult volunteers participated (written informed consent obtained). For both Italian and Dutch, 26 listeners were tested (16 females and 10 males per language, aged 21–48 for the Italians and 17–37 for the Dutch). There were 30 Romanian listeners (21 females and 9 males, aged 21–58). Participants received oral instructions, and the purpose of the study was also explained on an information sheet.

Listeners were tested individually in a quiet room at Leiden University (Dutch), Ca' Foscari University of Venice (Italian), or the University of Bucharest (Romanian). Stimuli were presented over headphones at a comfortable listening level. After each stimulus presentation, the participant was asked to respond which member of the lexical stress pair had been perceived. A response was given by clicking one of two on-screen response buttons that each showed one of the word forms, and the test only continued after a response had been given. Listeners were thus asked to respond even when they were not completely sure. An 8-stimulus practice session with stimuli taken from the main experiment was presented first. The main experiment contained 52 stimuli that were each presented once, in a different random order per listener. The interval between giving a response and presentation of the next stimulus was two seconds. The task lasted about six minutes.

2.4 Analysis

Per manipulated stimulus, listener responses were compared with responses to the corresponding baseline stimulus in paired Wilcoxon signed ranks tests. This was done to investigate if shifts in perception occurred as a result of manipulation. A one-tailed approach was taken, as changes to baseline responses could only be in one direction. Because of multiple comparisons per baseline stimulus, α was corrected to 0.01 (0.05/5). Chi-square analyses were used to establish whether

responses predominantly fell into one of two response categories, or were undetermined between the two categories.

For the sake of brevity, only results of particular stimuli are discussed: (1) [+Focus] words, as these contained the strongest acoustic cues, and (2) stylized pitch contours overlain on non-focused, duration-neutralized recordings, as these most closely matched the stimuli used in the test battery.

3. Results

The expectations were that responses to the baseline stimuli would be consistent within a language group and as intended. However, 5 out of 30 Romanian listeners responded to the baseline stimuli as either '1' (2×) or '2' (3×). We assumed that these listeners had difficulty with the task and they were therefore excluded from analysis.

An overview of the listeners' responses to the [+Focus] stimuli is given in Table 2.

Table 2. Percentages of perceived words per stimulus type and language background

	Dutch		Italian		Romanian	
	'canon	ka'nnon	'papa	pa'pa	'copii	co'pii
<i>Originals</i>						
S1F	100	0	100	0	52	48
S2F	0	100	0	100	20	80
<i>Duration only</i>						
S1F	89	11	96	4	76	24
S2F	11	89	0	100	48	52
<i>Pitch only</i>						
S1F	100	0	100	0	21	79
S2F	23	77	4	96	42	58
<i>Conflicting information</i>						
S1F with S2F duration	46	54	38	62	24	76
S2F with S1F duration	11	89	15	85	20	80
S1F with S2F pitch	19	81	89	11	52	48
S2F with S1F pitch	31	69	0	100	12	88

3.1 Baseline stimuli

Listeners responded as expected, except in the case of the S1F stimulus for Romanian listeners: half of the listeners perceived it as 2nd-syllable stress instead of 1st-syllable stress. Therefore, manipulations of this stimulus must be interpreted with caution.

3.2 Single cue: duration only

With only the duration cue available, Italian and Dutch listeners responded similarly as to the original stimuli. Consistent categorization was confirmed ($\chi^2 = 15.4$, $p < .001$). The Romanian listeners' responses to none of the stimuli changed significantly, but seemed to become somewhat less consistent to S2F. Without pitch information, S1F was classified as having stress on the 1st-syllable ($\chi^2 = 6.8$, $p = .009$), whereas classification was inconsistent on the other stimulus.

3.3 Single cue: F0 only

For the Dutch, responses did not differ significantly from the baseline, and classifications were consistent ($\chi^2 > 7.5$, $p \leq .006$). Comparable results were found for the Italian listeners ($\chi^2 > 22.2$, $p < .001$). Pitch-only stimuli with original pitch values both resulted in inconsistent classification by the Romanians ($\chi^2 < 3.2$, $p \geq .072$).

3.4 Conflicting cue stimuli

3.4.1 *Pitch replacements*

The Dutch listeners showed different responses as opposed to the originals when pitch had been replaced ($Z = -2.8$, $p \leq .005$). Stimulus S2F_pitchS1F was not classified consistently, whereas categorization of S1F_pitchS2F had changed ($\chi^2 = 9.8$, $p = .002$).

The Italian listeners showed no significantly different responses relative to the originals. All stimuli were consistently classified as either 1st- or 2nd-syllable stress ($\chi^2 = 15.4$, $p < .001$).

The Romanian listeners did not show responses significantly different from the originals. Listeners only consistently categorized stimulus S2F_pitchS1F ($\chi^2 = 14.4$, $p < .001$), perceiving it as having 2nd-syllable stress.

3.4.2 *Duration replacements*

For the Dutch listeners, significantly different responses were observed for the S1F stimulus as opposed to its original ($Z = -3.5$, $p = .001$), but not for S2F.

Categorization did not change to the other category as a result of duration manipulation; it either remained the same (S2F-manipulation, $\chi^2 = 15.4$, $p < .001$) or became undetermined (S1F-manipulation: $\chi^2 = 0.2$, n.s.).

For Italian listeners responses to duration-manipulated stimuli were different for the S1F-manipulation ($Z = -4.0$, $p < .001$), but not for the S2F-manipulation. Listeners were unsure how to classify S1F with a duration manipulation.

The Romanian listeners perceived both stimuli as carrying 2nd-syllable stress ($\chi^2 > 6.8$, $p \leq 0.009$).

3.5 Stylized pitch stimuli

The responses in case of stylized pitch stimuli are given in Table 3. The Dutch listeners classified all stylized stimuli as expected ($\chi^2 \Rightarrow 15.4$, $p \leq .006$). The Italians consistently classified all stylized pitch movements onto the first syllable ($\chi^2 \Rightarrow 11.6$, $p \leq .001$), but did so only for pitch excursions larger than 2 semitones on the second syllable ($\chi^2 \Rightarrow 3.8$, $p \leq .050$). Romanian listeners classified all stylized stimuli as expected ($\chi^2 \Rightarrow 9.0$, $p \leq .003$), but a 2nd-syllable 2 ST excursion ($\chi^2 = 3.2$, $p = .072$).

Table 3. Percentages of perceived words per stimulus type and language background

Accent location	Size (ST)	Dutch		Italian		Romanian	
		'canon	ka'non	'papa	pa'pa	'copii	co'pii
1st syllable	2	96	4	88	12	84	16
	4	96	4	92	8	88	12
	6	98	2	92	8	94	6
2nd syllable	2	13	87	50	50	22	78
	4	20	80	21	79	14	86
	6	15	85	17	83	10	90

4. Discussion

The first question was how F_0 and duration trade off as cues to prominence perception in three languages, and especially how listeners respond to natural F_0 changes when duration is neutralized. If listeners consistently perform a task based on stimuli of the latter type, we feel this validates our choice for this task as part of the test battery. The second question was if listeners from different language backgrounds would respond comparably to the F_0 manipulations used in

the test battery as to natural, native pitch accents. This would support our choice of a pseudo-language with stylized accents.

The *single cue* results showed that duration and F_0 both were a sufficient cue to lexical stress position in Dutch and Italian (van Katwijk 1974; Bertinetto 1980). For Romanian, it is more difficult to draw conclusions, as listeners as a group showed more variation in their responses, and it is — as yet — unclear what caused the inconsistent classification of one of the original stimuli; an informal pilot test had yielded consistent responses from 10 listeners. Still, our results suggest that neither F_0 nor duration is a sufficient cue to prominence location; the two cues interact and/or there is at least one other cue involved that remained audible in our materials.

Conflicting cue stimuli revealed that F_0 manipulation caused a category change in Dutch, but not in Italian (and Romanian) listeners. Reversely, duration manipulation in stimuli with intact F_0 caused comparable changes in categorization in Dutch and Italian. In sum, Dutch and Italian listeners respond comparably in single-cue cases, but may vary in how they process interactions of the two cues. This supports our assumption that identification of prominence location through F_0 alone is a feasible task as part of the test battery.

However, the exact pitch movements that cue prominence vary between languages. The second question therefore concerned how listeners from different language backgrounds would respond to stylized F_0 manipulations. We found that listeners generally responded the same to these stimuli; pitch accents on the first syllable yielded a clear majority of 1st-syllable responses, and pitch accents on the second syllable yielded a majority of 2nd-syllable responses. Only small pitch accents on the second syllable were not well recognized as such by Italian listeners, and the same stimulus category gave most inconsistency for the Romanians. However, taking into account that meaningful pitch accents need to be at least 1.5 to 3 ST (Gussenhoven & Rietveld 1985), the two larger accents may be considered the most important basis for comparisons.

So, can speech pitch perception be measured language-independently? The answer is yes, if test results are taken as a reflection of acoustic cue, i.e. F_0 , processing in a speech setting. We obtained comparable results with native language stimuli as with pseudo-language stimuli. But the answer is no, if test results are taken as a direct reflection of native language processing. Computing a JND for change in F_0 in speech-like stimuli is not the same as prominence perception in native speech communication; languages differ in how F_0 and other acoustic cues together impact interpretation.

Acknowledgements

This work was supported by the EU FP7-SME-222291 Dual Pro. We thank two anonymous reviewers for their helpful suggestions.

References

- Alfano, Iolanda. 2006. "La percezione dell'accento lessicale: un test sull'italiano a confronto con lo spagnolo". *Proceedings 2nd AISV* ed. by Renata Savy and Claudia Crocco, 632–656. Torriana: EDK Editore.
- Avesani, Cinzia, Mario Vayra & Claudio Zmarich. 2007. "On the articulatory bases of prominence in Italian". *Proceedings ICPHS XVI*. 981–984.
- Avram, Andrei. 1966. "Durata vocalelor și perceperea accentului în limba română [= The duration of vowels and the perception of stress in Romanian]". *Studii și cercetări lingvistice XVIII* 3.263–269.
- Avram, Andrei. 1970. "Sur le rôle de la fréquence dans la perception de l'accent en roumain". *Proceedings ICPHS VI, Prague 1967*. 137–139.
- Bertinetto, Pier M. 1980. "The perception of stress by Italian speakers". *Journal of Phonetics* 8.385–395.
- Bertinetto, Pier M. 1981. *Strutture prosodiche dell'italiano. Accento, quantità, sillaba, giuntura, fondamenti metrici*. Florence: Accademia della Crusca.
- Bertinetto, Pier M. & Michele Loporcaro. 2005. "The sound pattern of Standard Italian, as compared with the varieties spoken in Florence, Milan and Rome". *Journal of the International Phonetic Association* 35.131–151.
- Eefting, Wieke. 1991. "The effect of "information value" and "accentuation" on the duration of Dutch words, syllables, and segments". *Journal of the Acoustical Society of America* 89.412–424.
- Farnetani, Edda & Claudio Zmarich. 1997. "Prominence patterns in Italian: An analysis of F0 and duration". *INT-97*. 115–118.
- Fry, D. B. 1958. "Experiments in the perception of stress". *Language and Speech* 1.126–152.
- Giurgiu, Mircea. 2008. "Exploring fine phonetic detail for Romanian test-to-speech synthesis". *Acta Technica Napocensis — Electronics and Telecommunications* 49.44–47.
- Gussenhoven, Carlos & Toni Rietveld. 1985. "On the relation between pitch excursion size and pitch prominence". *Journal of Phonetics* 13.299–308.
- Heeren, Willemijn, Martine Coene, Bart Vaerenberg, Andrei A. Avram, Anna Cardinaletti, Luca Del Bo, Alexandru Pascu, Francesca Volpato & Paul J. Govaerts. in press. "Development of the AŞE test battery for speech pitch perception assessment in cochlear implant users". *Cochlear Implants International*.
- van Katwijk, Albert. 1974. *Accentuation in Dutch*. Amsterdam/Assen: Van Gorcum.
- Manolescu, Alis, Daniel Olson & Marta Ortega-Llebaria. 2009. "Cues to contrastive focus in Romanian". *Phonetics and phonology. Interactions and interrelations* ed. by Marina Vigário, Sonia Fróta and Maria João Freitas, 71–90. Amsterdam: John Benjamins.
- Moulines, Eric & Francis Charpentier. 1990. "Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones". *Speech Communication* 9.453–467.

- Nespor, Marina & Irene Vogel. 1986. *Prosodic phonology*. Dordrecht: Foris Publications.
- Nooteboom, Sieb G. & Johanna G. Kruyt. 1987. "Accents, focus distribution, and the perceived distribution of given and new information: An experiment". *Journal of the Acoustical Society of America* 82.1512–1524.
- Rossi, Mario. 1998. "Intonation in Italian". *Intonation systems. A survey of twenty languages* ed. by Daniel Hirst and Albert Di Cristo, 219–238. Cambridge: Cambridge University Press.
- Sluijter, Agaath M. C. & Vincent J. van Heuven. 1996. "Spectral balance as an acoustic correlate of linguistics stress". *Journal of the Acoustical Society of America* 100.2471–2485.
- Sluijter, Agaath M. C., Vincent J. van Heuven & Jos J. A. Pacilly. 1996. "Spectral balance as a cue in the perception of linguistic stress". *Journal of the Acoustical Society of America* 101.503–513.

Author's address

Willemijn Heeren
Leiden University Centre for Linguistics / Phonetics Lab
Cleveringaplaats 1
2311 BD Leiden
w.f.l.heeren@hum.leidenuniv.nl