MODAL VERSUS CREAKY FILLER PARTICLES IN ROMANIAN CONNECTED SPEECH

OANA NICULESCU[1]

Abstract. This paper represents a preliminary acoustic analysis of filler particles in terms of voice quality (i.e., modal vs. creaky phonation). The main research questions addressed in this study revolve around which particular voice parameters are indicative of (non)modal phonation of fillers used by healthy speakers of Standard Romanian, and whether the function of the filler particle varies with different voice qualities. The analysis is carried out on Romanian connected speech data extracted from the Ro-Phon corpus (non-pathological speech), an open-access linguistic resource developed during our postdoctoral research project financed by UEFISCDI (2020 – 2022).

Keywords: filler particles, voice quality, creaky voice, spontaneous speech, Ro-Phon corpus, Romanian data.

1. INTRODUCTION

Phonetic studies of disfluency phenomena based on Romanian data are still in the early stages. Even less considered is the link between voice quality (i.e., phonation type) and the various forms and functions of filler particles in spontaneous speech. Consequently, the purpose of this study is to investigate, from a perspective of non-pathological voice quality, the acoustic correlates of modal and creaky filler particles in Romanian connected speech. Our main research questions revolve around which particular voice parameters are indicative of (non)modal phonation of fillers used by healthy speakers of Standard Romanian, and whether the function of the filler particle varies with different voice qualities. In order to address these questions we rely on data extracted from Ro-Phon corpus (Niculescu 2021), an open-access linguistic resource developed as part of our postdoctoral research project financed by UEFISCDI (2020 – 2022).

In a previous article (Niculescu 2023), we adhered to a terminology and definition put forward by Belz (2023), where "a phonetic exponent which is segmentally structured, semantically empty, syntactically unconstrained, and does not show an interjectional function is classified as a filler particle". Within this framework and based on data from the Ro-Phon corpus, we classified filler particles according to (i) timbre and structure into *vocalic* (mid-open or closed central vowels present within the Romanian vocalic system), *vocalic-nasal* (central mid-open or closed vowel with a nasal coda) and *nasal* (bilabial or dental) filler particles;

[1] Institute of Linguistics "Iorgu Iordan – Alexandru Rosetti" of the Romanian Academy, oeniculescu@yahoo.com.

(ii) position in relation to the adjacent silent pause, distinguishing between *pre-pausal* (e.g. zile‿@m <pauză>) *post-pausal* (<pauză> @ăm: atestatu(l)) *inter-pausal* (<pauză> @ă: <pauză>) and *concatenated* filler particles (e.g. cunoscut‿@ăm‿+un alt‿@ă‿un). These categories are illustrated in the following examples:

(1) cam de: <pauză> aproape doi ani de zile‿**@m** <pauză> **@ă:** <pauză> plănuiam să mai am un copil (*Ro-Phon*, S_1)
‘for about <pause> nearly two years *m* [m] <pause> *ă* [ə] <pause> I was planning to have another child’

(2) având deja toate-acele trăiri intense <pauză> **@ă**‿reușes(c)‿+să: să fiu mai‿**@ăm** <pauză> nu ș(ti)u dacă chiar dimplomată (*Ro-Phon*, S_1)
‘already having all of those intense feelings <pause> *ă* [ə] I am able to to be more *ăm* [əm] <pause> I donno if necessarily more diplomatic’

(3) unul fiind chiar **@ă** coleg +de: de clasă <pauză> așa și altu(l) prin el am cunoscut‿**@ăm**‿+un alt‿**@ă**‿un al(t) prieten (*Ro-Phon*, S_6)
‘one of them actually being *ă* [ə] colleague of of class <pause> so and another through him I met *ăm* [əm] another *ă* [ə] another friend’

(4) este cea legată de: <pauză> **@ă**:‿atestatu(l) de instructor <pauză> **@ăm:** atestatu(l) de instructor‿**+@ă:**‿însem^-% înseamnă (*Ro-Phon*, S_6)
‘is the one related to <pause> *ă* [ə] the instructor certificate <pause> *ăm* [əm] the instructor certificate *ă* [ə] me- means’

The research on creaky fillers is not as extensive as that on modal voice outputs. Most of the studies, which primarily focus on Germanic languages, examine either the production or perception of filler particles in modal phonation both in controlled and spontaneous speech. In Swedish data however, based on the SweDia 2000 interview material, Horne (2006, 2009) identified three phonetic realizations of “EH”, with the vocalic filler particle in modal phonation having the highest frequency (n = 61), followed by the nasalized vowel or vocalic-nasal filler particle in modal voicing (n = 43), while the creaky filler EH had the fewest occurrences (n = 31). As to the function played by non-modal voice fillers in the discourse, Horne (2009) noted that they signal an attitude of uncertainty/indecisiveness. Non-creaky fillers were found to be associated with a discourse segment boundary or to function as prominence markers, preceding focused words. Data coming from German (Belz 2017) reveal that besides the vocalic (*äh*) and vocalic-nasal filler particles (*ähm*), an additional form surfaces as a glottal variant. No distinctive functions of this creaky filler were further discussed.

The remaining part the article is organized as follows. Section 2 showcases the voice quality analysis carried out through the acoustic measurements provided by the Voice Report in Praat. The results are outlined in the third section, while discussions and conclusions are presented in the final section.

2. VOICE QUALITY ANALYSIS

Following Laver’s (1980) classification, the physiological portrayal of various voice qualities is carried out in connection with three parameters of muscular tension, namely: (i) the *adductive tension*, i.e, "the force by which the arytenoids are drawn together, so that the cartilaginous glottis is adducted", (ii) the *medial compression*, i.e., "the force by which the ligamental glottis is closed, through the approximation of the vocal processes of the arytenoids", and (iii) the *longitudinal tension*, i.e., "the tension of the vocal folds"(Gobl and Ni Chasaide 2010: 395). Against this background, modal voice is defined as the neutral type of phonation, with a moderate adductive tension, medial compression and longitudinal tension (Gobl and Ni Chasaide 2010: 399), whereas creaky voice entails a high adductive tension and medial compression, correlated with a lesser longitudinal tension (Gobl and Ni Chasaide 2010: 401).

The voice analysis carried out in this paper pertains only to filler particles in spontaneous speech. As such, our study is carried out on standard Romanian connected speech data extracted from the manually aligned Ro-Phon corpus (Niculescu 2021) containing addressed monologues pertaining to 12 monolingual young adults (6 female, 6 male), representative of the Southern dialect, without any speaking or hearing impairments (non-pathological speech). What is important to point out is that the corpus was not explicitly designed to elicit filler particles, nevertheless various disfluencies including fillers were manually annotated in the transcriptions (through TextGrids in Praat, first tier). The annotation used for filler particles is "@" placed in front of the segment, without differentiating between timbre, structure or voice quality. In light of this, we would like emphasise the influential role an annotating system plays when analysing various phenomena in connected speech, especially when dealing with disfluencies. By making use of an inclusive as well as efficient transcription system (see, for instance, Hough et al. 2015), we can better understand the dynamic interaction between fluency and disfluency phenomena both in native and nonnative Romanian speech[2].

2.1. Parameters

For this articles, we will be using the measurements provided by the Voice Report in Praat, typically taken on sustained vowels. This report is generated through a command found in the *Pulses* menu (i.e., glottal pulses visible in the SoundEditor window as blue vertical lines distributed across the waveform). The user must first make the pulses visible by selecting Show pulses from the Pulses menu. Secondly, in order to get more precise measurements, an optimization procedure must pe performed by choosing "Optimize for voice analysis" (i.e., cross-correlation analysis method) from the "Pitch settings". We also changed the analysis settings from "Advanced Pitch settings" to "Very accurate". Additionally, we modified the default analysis configuration from the "Advanced Pitch settings" menu to "Very accurate".

The Voice Report in Praat provides information related to: (1) *pitch* – median, mean, standard deviation, minimum and maximum pitch; (2) *pulses* – number of pulses, number of periods, mean period, and standard deviation of period; (3) *voicing* – fraction of locally unvoiced frames (close to null when the segment is voiced), number and degree of voice breaks; (4) *jitter* – local, absolute, rap, ppq5, ddp[3]; (5) *shimmer* – local, dB, apq3, apq5,

[2] For Romanian L2 corpus, see the LECOR project (Mîrzea Vasile 2020, Barbu et al. 2023).

[3] For a detailed presentation of each measurement, see Praat Intro Manual, "Jitter".

apq11, dda[4]; and (6) *Harmonicity of the voiced parts only* – mean autocorrelation, mean noise-to-harmonics ratio, mean harmonics-to-noise ratio.

Apart from jitter and shimmer measures (related to *f*0 perturbation estimates), differences in voice quality have also been quantified by means of spectral tilt (i.e., harmonic amplitude measures) as well as by noise and periodicity measures such as the harmonics-to-noise ratio (HNR) and the cepstral peak prominence (CPP). An extensive body of literature has shown that the difference between the first and second harmonics and, particularly, the formant-corrected amplitude difference (H1*-H2*) correlates both with the Open Quotient (OQ) and the Contact Quotient (CQ) estimates from electroglottography (Keating *et al.* 2023), where a lower harmonic difference, especially between the first and second harmonic, denotes a greater glottal constriction, characteristic of non-modal phonation. Lower values for HNR correlate with creaky voice due to a decrease in periodicity, characteristic to this phonation type (Davidson 2019). Studies have shown that CPP (considered a measure of HNR) is a reliable and valid parameter of measuring voice quality, not only in sustained vowels (prototypical tokens for assessing voice phonation), but also in spontaneous speech data (Watts and Awan 2011). An additional acoustic measure of voice quality, although it received less attention in the literature (Kirk *et al. 1*993, and, recently, Ludusan *et al.* 2023), pertains to formant frequencies, in particular the first formant, with higher F1 values observed during creaky phonation (for a cross-linguistic overview, see Gordon and Ladefoged 2001: 400). A more comprehensive review of creaky voice is undertaken in the following section.

2.2. Types of creaky voice

In the literature, a prototypical creaky voice is defined as having three key properties, namely a lower fundamental frequency than modal voice (i.e., low rate of vocal fold vibration), an irregular *f*0 (random or multiply pulsed) and a constricted glottis (the vocal fold are close together) manifested by low glottal airflow due to a small peak glottal opening followed by a long closed stage denoting a low spectral-to-noise ratio (Keating et al. 2015, Keating et al. 2023). However, as Gobl and Ni Chasaide (2010: 401) eloquently summarised "although every voice quality" (i.e., modal, breathy, whispery, creaky, lax and tense voice, in line with Laver's (1980) classification) "varies dynamically in the course of an utterance, creaky voice is particularly variable". In light of this, Keating *et al.* (2015) identify and describe an additional five phonetic representations of creaky voice, such as:

- *vocal fry* → characterised by constricted glottis (correlated with a high EGG Contact Quotient), a low *f*0, without it necessarily being irregular (as in the case of prototypical creak), a low difference between the two harmonics (H1-H2) indicating a greater constriction (measurement closely related to the glottal Open Quotient), a relative high HNR alongside a damping of the glottal pulses.
- *multiply pulsed voice* → has a fundamental frequency composed of alternating longer and shorted pulses (i.e., irregular *f*0, leading to multiple-pulsed creak), with a high SHR (resulting from an increased number of subharmonics).
- *aperiodic voice* → considered another variant of *f*0 irregularity (extremely irregular *f*0), without it having low values, while the H1-H2 harmonics also register a low value correlated with higher glottal constriction.

[4] See Praat Intro Manual, "Shimmer", for an in-depth account of each measurement.

- *nonconstricted creak* → similar to prototypical creak, this subtype possess a low and irregular fundamental frequency, however H1-H2 is high denoting rather a spreading of the glottis and not a glottal constriction (correlated with a low CQ which indicates a small glottal constriction).
- *tense/pressed voice* → constricted glottis without a low nor an irregular fundamental frequency, having low H1-H2 and a high CQ.

So as to aid in visualising the acoustic range between modal and creaky phonation, in what follows we provide waveform-based examples of (non)creaky pre- and post-pausal vocalic (Figures 1–2) as well as nasal filler particles (Figures 3–4) extracted from both female and male monologues. Given the scarcity of Romanian data illustrations, these images can broaden our understanding not only in connection to the phonetic properties of filler particles in spoken language, but also in relation to the acoustic correlates of the modal-creaky voice continuum.

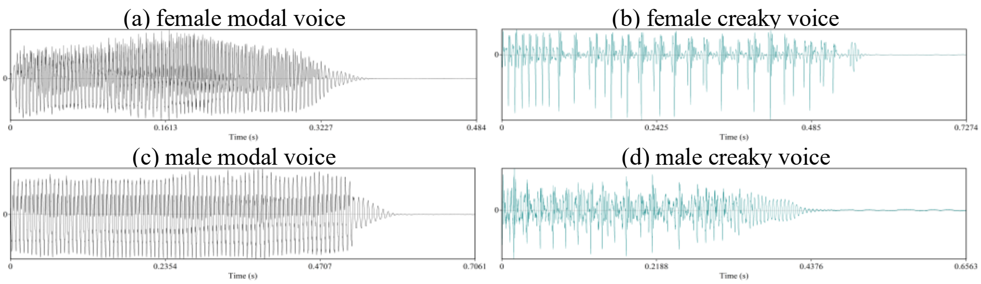

Figure 1
Waveform of pre-pausal vocalic filler particle in modal and creaky phonation

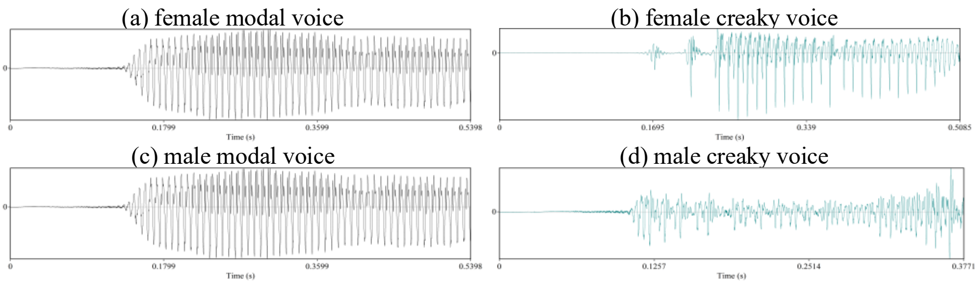

Figure 2
Waveform of post-pausal vocalic filler particle in modal and creaky phonation

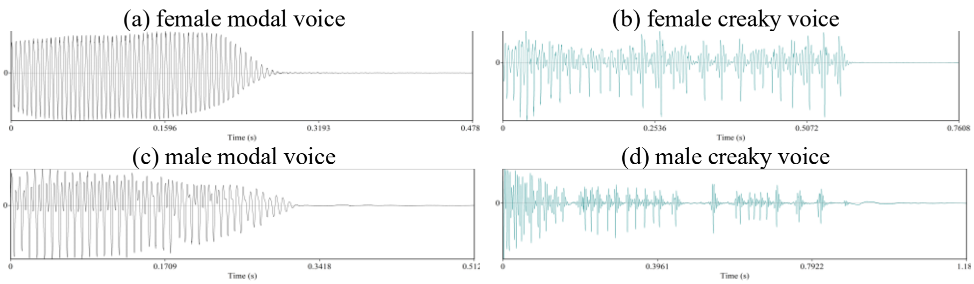

Figure 3
Waveform of pre-pausal nasal filler particle in modal and creaky phonation

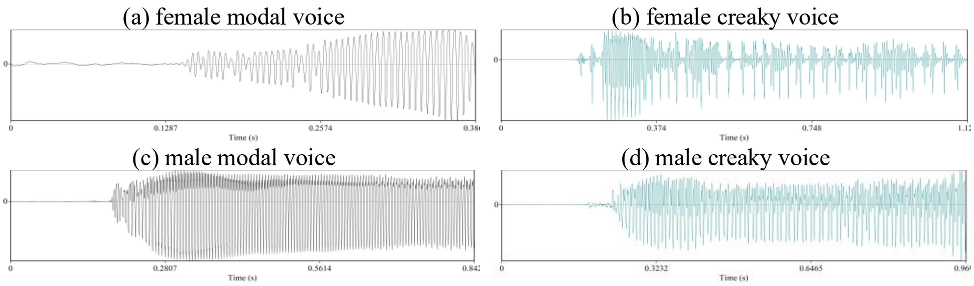

Figure 4

Waveform of post-pausal nasal filler particle in modal and creaky phonation

All modal tokens are similar in the sense that the harmonics are equally spaced from each other. With regard to non-modal voice, each spectral slice is different, showcasing a series of both periodic and aperiodic creaks. Apart from the nasal filler in male speech which illustrates a periodic creak, all other outputs display highly irregular spaced vocal pulses. Another important aspect worth mentioning is that creaky phonation in Romanian speech is not restricted to utterance-final position (as in American English Slifka 2006), nor does it surface exclusively in monologues pertaining to only one gender.

When measuring acoustic parameters of voice quality, it is important to differentiate between data obtained from healthy voices and data derived from pathological voices. Table 1 correlates non-pathological with pathological speech results reported from various high-profile studies.

Measurements	Healthy Voice			Pathological Voice	
	Teixeira and Fernandes (2014)		Tylečková and Skarnitzl (2019)	Guimarães (2007)	Teixeira et al. (2013)
	Male	Female			
Jitta (μs)	32,11 (18,028)	14,10 (8,674)	–	83.2	83.200
Jitt (%)	0,3619 (0,17637)	0,3324 (0,17168)	1.83 (1.97)	1.04	1.040
rap (%)	0,1762 (0,10883)	0,2014 (0,20217)	–	0.68	0.680
Ppq5 (%)	0,2111 (0,10490)	0,1965 (0,12613)	–	–	0.840
Shim (%)	2,2873 (1,26242)	2,7458 (2,31167)	13.02 (6.75)	3.81	3.810
ShdB (dB)	0,2032 (0,11355)	0,2389 (0,19675)	–	0.35	0.350
apq3 (%)	1,1429 (0,69601)	1,3516 (1,13736)	–	–	–
apq5 (%)	1,4286 (0,78708)	1,6614 (1,51868)	–	–	–
HNR (dB)	24,0095 (4,36920)	24,9474 (4,48382)	9.4 (4.05)	–	7
CPP	–	–	20.3 (3.69)	–	–

Table 1.

Threshold values for healthy and pathological voice[5]

[5] The jitter and shimmer measurement are made by MDVP (i.e., Multidimensional Voice Program) or VoiceSauce, having the following equivalents in Praat: Jitta (μs) = Jitter (local, absolute), Jitt (%) = Jitter (local), rap (%) = Jitter (rap), Ppq5 (%) = Jitter (ppq5), Shim (%) = Shimmer (local), ShdB (dB) = Shimmer (local, dB), apq3 (%) = Shimmer (apq3), and apq5 (%) = Shimmer (apq5).

The non-pathological speech results are reported by Teixeira and Fernandes (2014) with values extracted from sustained vowels /a, i, u/, at High, Low and Neutral tones, belonging to 34 female and 7 male healthy speakers from the Saarbrücken Voice Database. Tylečková and Skarnitzl (2019) also contributed to the non-pathological speech results by presenting the values extracted from the Database of Common Czech on sustained open vowels /a a:/ belonging to fifty healthy male speakers. The pathological speech results represent the most recurrent values cited in the literature as a threshold for pathology.

As a phonation type, creaky voice has received attention from various research fields. Going beyond clinical settings and forensic phonetics, creaky phonation has been shown to have different linguistic and extralinguistic functions. For instance, there are languages which have a phonemic contrast between creaky and modal voice. Chadic languages contrast between modal and creaky voice obstruents (Ladefoged and Maddieson 1996). Jalapa Mazatec even has a three-way phonation contrast, differentiating between modal voiced, breathy voiced, and creaky voiced vowels (Gordon and Ladefoged 2001: 387, in line with Ladefoged and Maddieson 1996), similar to San Lucas Quiaviní Zapotec (Munro and Lopez 1999 *apud* Gordon and Ladefoged 2001: 398).

In a recent article, Lisa Davidson (2021) explores the various functions creaky voice fulfils cross-linguistically, ranging from a segmental or suprasegmental element, creating lexical contrast, to a prosodic feature, signalling turn-taking or end of a phrase. The review article also illustrates how creaky voice can function as a sociolinguistic marker, conveying speaker attitudes such as irritation (Vietnamese) or indecisiveness in the context of filler particles (Swedish). Furthermore, there are increasingly more studies that use creaky voice as a feature for pitch detection algorithms and recognition systems, automatic speech processing as well as voice synthesis.

In our acoustic assessment of creaky and non-creaky filler particles, we rely both on qualitative analysis (visual inspection of waveform and broadband spectrogram) as well as on quantitative assessment of the data performed by analysing the values extracted through the Voice Report in Praat. Based on the information provided thus far regarding the acoustic properties of modal and creaky phonation, we expect to find a lower f_0 associated with higher levels of jitter and shimmer surfacing in non-modal fillers. The modal voice is predicted to have a higher HNR, while creaky fillers are expected to have a higher degree of glottal constriction as indicated by spectral tilt measurements and CPP.

3. RESULTS

This section presents the results of our acoustic analysis based on the estimated values of the corresponding voice parameters outlined in the previous section. Subsequent sections are organised as following: modal *vs.* creaky phonation of post-pausal vocalic filler particle in female monologues (section 3.1), modal *vs.* creaky phonation of pre-pausal vocalic-nasal filler particle in female speech (section 3.2), and modal *vs.* creaky phonation of pre-pausal nasal filler particle in male speech (section 3.3).

For each account, two waveforms and broadband spectrograms of the voice qualities under study are displayed in conjunction with the measurements summarised by the Voice Report in Praat (see section 2.1). CPP values were manually obtained, while H1-H2 measurements were extracted through a script wrote by C. DiCanio (https://www.acsu.buffalo.edu/~cdicanio/scripts.html).

3.1. Vocalic Filler Particle

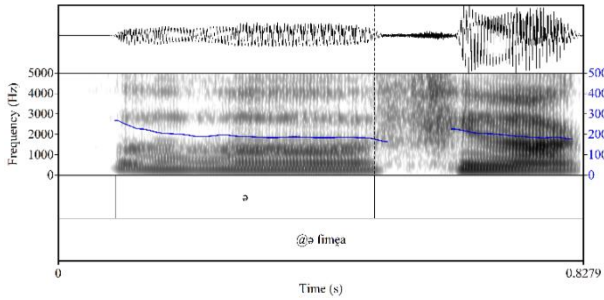

Figure 5
Female modal phonation [# ə fimẹa] *(Ro-Phon, S_1)*
Waveform and broadband spectrogram of post-pausal vocalic filler particle

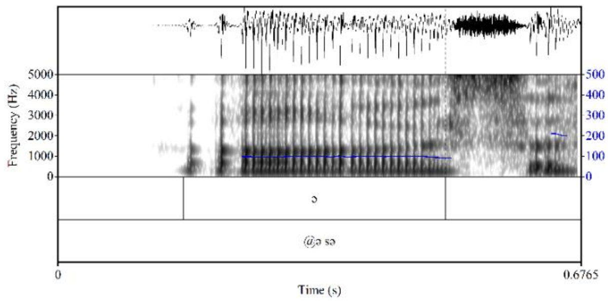

Figure 6
Female creaky phonation [# ə sə] *(Ro-Phon, S_9)*
Waveform and broadband spectrogram of post-pausal vocalic filler particle

Figure 5 illustrates the vocalic filler in modal phonation, with an overall duration of 545ms, present in the bordered portion of the utterance “nu am suficien(t) de mul(t) timp liber să mi-l petrec cu: +fi-mea <pauză> @ă:‿fi-mea care se plânge la maică-mea to(t) timpu(l) că io nu mă joc cu ea” ‘I do not have enough free time to spend with my daughter <pause> *ă* [ə] my daughter who complains to my mother all the time that I do not play with her’, while Figure 6 showcases the filler in creaky phonation (349ms) present in the bordered portion of the utterance “ci să și-(î)mi fac prieteni în zona-(î)n care locuiesc și <pauză> @ă:‿să fie așa o petrecere continuă aproape tot anu(l)” ‘but likewise to make friends in the area I live in and <pause> *ă* [ə] to be like a never-ending part nearly all year round’. The non-creaky filler surfaces in the editing phase of an identical repetition, fulfilling a discourse planning function, while the creaky filler appears at a deeper discourse boundary, signalling a delay, or better said, “embodying” the delay (Schegloff 2010: 141).

In terms of acoustic measurements, we observe a regular f_0 tracked throughout the entire vocalic segment, apart from the two glottalizations at the beginning of the creaky filler, having a higher average frequency in the case of modal phonation 178 Hz (SD = 6) vs 161 Hz (SD = 54) in creaky voice. We also notice that modal phonation has a lower standard deviation of the fundamental frequency.

Table 2

Modal versus creaky phonation of vocalic filler particle
Results from the Voice Report in Praat

Measurements	Modal Voice	Creaky Voice
1. Duration	0.545015 s	0.349076 s
2. Pitch		
2.1 Median pitch	176.783 Hz	198.686 Hz
2.2 Mean pitch	177.830 Hz	160.833 Hz
2.3 Standard deviation	6.492 Hz	54.027 Hz
2.4 Minimum pitch	154.226 Hz	91.133 Hz
2.5 Maximum pitch	188.956 Hz	218.428 Hz
3. Pulses		
3.1. Number of pulses:	97	35
3.2. Number of periods:	96	33
3.3 Mean period	5.612812E-3 s	6.403973E-3 s
3.4 Standard deviation of period	0.193396E-3 s	2.331983E-3 s
4. Voicing		
4.1 Fraction of locally unvoiced frames	0 (0 / 164)	17.143% (18 / 105)
4.2 Number of voice breaks	0	1
4.3 Degree of voice breaks	0 (0 s / 0 s)	38.373% (0.133952 s / 0.349076 s)
5. Jitter		
5.1 Jitter (local)	0.439%	7.061%
5.2 Jitter (local, absolute)	24.668E-6 s	452.209E-6 s
5.3 Jitter (rap)	0.193%	3.883%
5.4 Jitter (ppq5)	0.208%	6.058%
5.5 Jitter (ddp)	0.580%	11.649%
6. Shimmer		
6.1 Shimmer (local)	1.828%	18.265%
6.2 Shimmer (local, dB)	0.171 dB	1.588 dB
6.3 Shimmer (apq3):	0.635%	11.967%
6.4 Shimmer (apq5)	0.658%	9.760%
6.5 Shimmer (apq11):	1.240%	--undefined--
6.6 Shimmer (dda):	1.904%	35.901%
7. Harmonicity of the voiced parts only		
7.1 Mean autocorrelation	0.987220	0.764026
7.2 Mean noise-to-harmonics ratio	0.021833	0.358235
7.3 Mean harmonics-to-noise ratio	26.999 dB	6.383 dB

Creaky phonation has a lower number of pulses and periods compared to modal voicing, as well as a higher degree of devoicing based on the results given by the fractions of locally unvoiced frames. As expected, higher values of jitter and shimmer were found in the case of vocalic creaky filler particles (7% *vs.* 0.4%, and 18.2% *vs.* 1.8% respectively). Moreover, independent of phonation type, both measured jitter and shimmer values are above the threshold for pathological voices (see Table 1). These values are to be expected when working on connected speech data. The Praat algorithm returning an “undefined” value, e.g. shimmer (apq11), occurs when there are aperiodic glottal pulses.

As mentioned previously, the difference between amplitudes of the first and the second harmonic (H1-H2) is a measure of vocal fold constriction, with a lower value indicating a higher glottal constriction. This measurement is closely related to cepstral peak prominence (CPP) as well as to the harmonics-to-noise ratio (HNR). What we remark in the case of non-modal filler particle is actually a higher degree of glottal constriction indicated both by the spectral tilt measure (-10.7 Hz *vs.* 12.4 Hz) as well as by noise-related measures such as CPP (4.9 dB *vs.* 15.1 dB) and HNR (6.3 dB *vs.* 26.9 dB).

3.2. Vocalic-Nasal Filler Particle

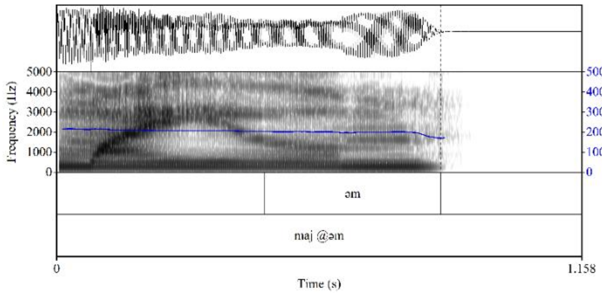

Figure 7
Female modal phonation [maj əm #] *(Ro-Phon, S_5)*
Waveform and broadband spectrogram of pre-pausal vocalic-nasal filler particle

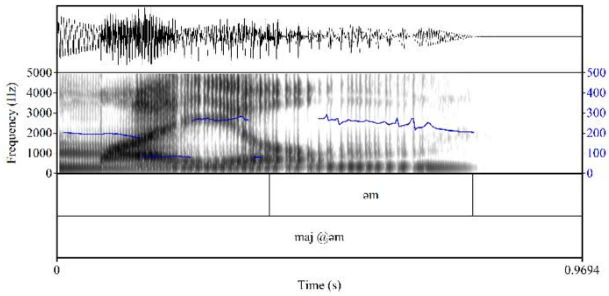

Figure 8
Female creaky phonation [maj əm #] *(Ro-Phon, S_5)*
Waveform and broadband spectrogram of pre-pausal vocalic-nasal filler particle

The modal vocalic-nasal filler particle is depicted in figure 7 in the bordered portion of the utterance: "și chiar îmi dores(c)‿să mai fac chestii spontane da(r) o să vedem o să +mai‿@ăm: <pauză> mai vedem pe parcurs" 'and I really want to do more spontaneous stuff but we will see we will *ăm* [əm] <pause> we will see along the way'. Surfacing within a repetition, the modal voice filler functions as an initiation of a repair.

The creaky vocalic-nasal filler particle is presented in figure 8, being visible in the bordered portion of the utterance: "apropo că vă spuneam io cu:‿@m <pauză> că sînt așa mai‿@ăm: <pauză> mă rog <pauză> dificilă^ sau mă rog mai pretențioasă^" 'by the way I was telling you about *m* [m] <pause> that I am more like *ăm* [əm] <pause> whatever <pause> difficult or whatever more pretentious'. The filler in this context conveys an attitude of uncertainty, emphasised by neighbouring discourse markers.

Both filler particles share a similar duration (389 ms modal *vs.* 375 ms creaky). However, unlike the previous example, where the creak was restricted to the vowel segment only, here it originates in the previous glide spreading throughout the vocalic-nasal filler particle. There is a steady flat f_0 tracked across the entire modal filler, whereas a high, irregular pitch characterises the creaky filler. The fundamental frequency peaks at 203 Hz for modal phonation and at 293 Hz for non-modal voicing of the filler particle.

Table 3

Modal versus creaky phonation of vocalic-nasal filler particle
Results from the Voice Report in Praat

Measurements	Modal Voice	Creaky Voice
1. Duration	0.388965 s	0.375359 s
2. Pitch		
2.1 Median pitch	200.083 Hz	248.975 Hz
2.2 Mean pitch	197.493 Hz	243.571 Hz
2.3 Standard deviation	7.698 Hz	22.107 Hz
2.4 Minimum pitch	170.132 Hz	204.385 Hz
2.5 Maximum pitch	202.871 Hz	293.333 Hz
3. Pulses		
3.1. Number of pulses:	77	66
3.2. Number of periods:	76	62
3.3 Mean period	5.070878E-3 s	4.109889E-3 s
3.4 Standard deviation of period	0.220862E-3 s	0.452052E-3 s
4. Voicing		
4.1 Fraction of locally unvoiced frames	0 (0 / 117)	16.071% (18 / 112)
4.2 Number of voice breaks	0	0
4.3 Degree of voice breaks	0 (0 s / 0 s)	0 (0 s / 0 s)
5. Jitter		
5.1 Jitter (local)	0.728%	5.152%
5.2 Jitter (local, absolute)	36.913E-6 s	211.746E-6 s
5.3 Jitter (rap)	0.304%	2.819%
5.4 Jitter (ppq5)	0.356%	3.104%
5.5 Jitter (ddp)	0.913%	8.457%
6. Shimmer		
6.1 Shimmer (local)	3.118%	19.759%
6.2 Shimmer (local, dB)	0.361 dB	1.421 dB
6.3 Shimmer (apq3):	1.016%	1.979%
6.4 Shimmer (apq5)	1.240%	1.412%
6.5 Shimmer (apq11):	2.235%	1.145%
6.6 Shimmer (dda):	3.048%	5.938%
7. Harmonicity of the voiced parts only		
7.1 Mean autocorrelation	0.995367	0.797830
7.2 Mean noise-to-harmonics ratio	0.004684	0.325061
7.3 Mean harmonics-to-noise ratio	25.976 dB	9.752 dB

The fully creaky filler particle displays a lower number of pulses and periods compared to the vocalic-nasal filler produced in modal voice. Simmilar to the previous example, there is partial devoicing in the non-modal filler as well as higher rates of jitter and shimmer (5.1% *vs* 0.7%, and 19.7% *vs* 3.1%). Both f_0 perturbation measures are above the 1.04% (jitter) and 3.81% (shimmer) stated limits for detecting voice pathologies. There are no voice breaks identified in either phonation type. Equally significant is that in both type of fillers analysed so far on female speech, the first formant average frequency values are higher during creaky phonation (higher F1 correlates with a lower vocal height): 500 Hz *vs.* 385 Hz in the case of vocalic filler, and 457 Hz *vs.* 374 Hz for the vocalic-nasal filler.

Regarding the difference between amplitudes of the first and the second harmonic, we find a highly unusual result (although a similar situation was also found by Davidson 2019) where there is a positive H1-H2 in creaky phonation (21.0 Hz), significantly higher than the value obtained in modal voicing (9.8 Hz). Since the analysis in carried out on spontaneous data and the results are extracted from only one token, for the moment, we cannot make any further inferences. Such cases will be investigated in future studies.

As reported by noise-related measures such as CPP (8.5 dB *vs.* 13.1 dB) and HNR (9.7 dB *vs.* 25.9 dB), the degree of glottalization is higher in the non-modal filler particle.

3.3. Nasal Filler Particle

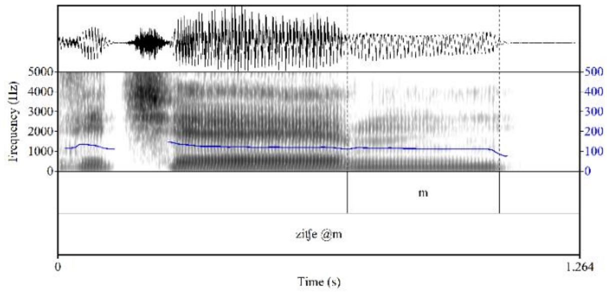

Figure 9
Male modal phonation [ziʧe m #] *(Ro-Phon, S_6)*
Waveform and broadband spectrogram of pre-pausal nasal filler particle

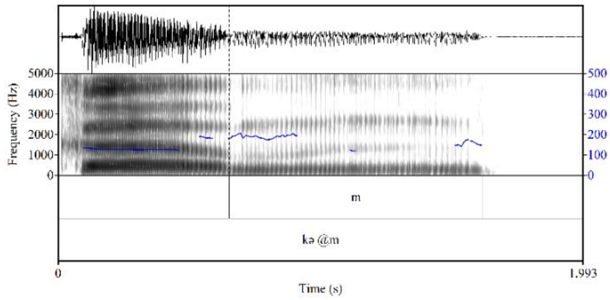

Figure 10
Male creaky phonation [kə m#] *(Ro-Phon, S_8)*
Waveform and broadband spectrogram of pre-pausal nasal filler particle

Figure 9 illustrates the non-creaky nasal filler particle from the bordered portion of the utterance "din asta cum îi zice:‿@m: <pauză> s(ă) (z)ic un fel de malaxor" 'one of this how is it called *m* [m] <pause> let's say a type of belnder', while figure 10 depicts the creaky nasal filler derived from the bordered portion of the utterance "și nu era clar că:‿@m: <pauză> @ă:m:‿ăla dinainte: a(d)ică‿ăla pleacă: vine ăsta adică nu se știa" 'and it was not clear that *m* [m] <pause> *ăm* [əm] the one before meaning that one is leaving the other one is coming that is no one knew'. The non-creaky filler particle fulfils a distinctive word-selection/word-retrieval function in the discourse. In contrast, the nasal filler particle in male creaky phonation signals uncertainty/indecisiveness (observation in line Horne 2009), function emphasised by the extended cluster of filler particles and silent pauses.

With regard to overall duration, the nasal filler is almost three times longer in creaky voicing than in modal phonation (962 ms creak *vs.* 369 ms modal).The nasal filler in modal phonation has a flat f_0 contour with an average frequency of 113 Hz (SD = 4.6), peaking at 120 Hz. Upon closer inspection of the spectrogram, we observe that f_0 is tracked only at the edges of the non-modal filler, displaying a highly irregular pattern, with creak originating in the final part of the preceding vowel and spreading throughout the pre-pausal filler particle.

Table 4

Modal versus creaky phonation of nasal filler particle
Results from the Voice Report in Praat

Measurements	Modal Voice	Creaky Voice
1. Duration	0.368960 s	0.962433 s
2. Pitch		
2.1 Median pitch	113.143 Hz	123.065 Hz
2.2 Mean pitch	113.110 Hz	133.120 Hz
2.3 Standard deviation	4.677 Hz	39.900 Hz
2.4 Minimum pitch	85.272 Hz	87.234 Hz
2.5 Maximum pitch	119.919 Hz	224.981 Hz
3. Pulses		
3.1. Number of pulses:	42	52
3.2. Number of periods:	41	40
3.3 Mean period	8.829835E-3 s	7.279351E-3 s
3.4 Standard deviation of period	0.367558E-3 s	2.282919E-3 s
4. Voicing		
4.1 Fraction of locally unvoiced frames	0 (0 / 110)	20.415% (59 / 289)
4.2 Number of voice breaks	0	6
4.3 Degree of voice breaks	0 (0 s / 0 s)	63.407% (0.610251 s / 0.962433 s)
5. Jitter		
5.1 Jitter (local)	1.187%	8.935%
5.2 Jitter (local, absolute)	104.808E-6 s	650.427E-6 s
5.3 Jitter (rap)	0.412%	4.439%
5.4 Jitter (ppq5)	0.515%	4.325%
5.5 Jitter (ddp)	1.236%	13.317%
6. Shimmer		
6.1 Shimmer (local)	3.349%	20.254%
6.2 Shimmer (local, dB)	0.302 dB	1.753 dB
6.3 Shimmer (apq3):	1.064%	7.311%
6.4 Shimmer (apq5)	1.556%	--undefined--
6.5 Shimmer (apq11):	1.970%	--undefined--
6.6 Shimmer (dda):	3.191%	21.934%
7. Harmonicity of the voiced parts only		
7.1 Mean autocorrelation	0.993767	0.699869
7.2 Mean noise-to-harmonics ratio	0.006608	0.552631
7.3 Mean harmonics-to-noise ratio	27.745 dB	4.889 dB

When compared to previous cases of vocalic(nasal) female filler voicing, we observe that male phonation has a higher degree of voice brakes surfacing in creaky fillers (63.4% *vs.* 38.3% in the vocalic context and 0% within the vocalic-nasal output of female speech). We also notice a higher rate of devoicing (20.4%) resulting from the fraction of locally unvoiced frames (59 out of the 289 frames)

Both measurements of f_0 perturbation register a greater value in creaky voicing (8.9% jitter, 20.2% shimmer) as opposed to modal voicing (1.1% jitter, 3.3% shimmer). Due to highly aperiodic glottal pulses, the Praat algorithm returned "undefined" values for the five-point and eleven-point Amplitude Perturbation Quotient.

Similar to previous examples, glottal constriction is higher in the case of creaky nasal filler particle in comparison with modal voice filler phonation. This observation is corroborated by the difference between the first and second harmonics (-1.8 Hz *vs.* 9.8 Hz) as well as by the cepstral peak prominence (7.7 dB *vs.* 11.8 dB) and harmonic-to-noise ratio (4.8 dB *vs*. 27.7 dB). Even though the sonorant is strongly glottalized, it still carries perceivable nasal features.

4. DISCUSSIONS AND FINAL REMARKS

This paper represents a preliminary acoustic analysis of Romanian filler particles in terms of voice quality (i.e., modal *vs.* creaky phonation derived from speakers without any hearing or speaking impairments). This type of research could only be carried out on speech data recorded, aligned (either manually or automatically), and properly annotated to better suit analyses at the interface between phonetics and laboratory phonology, such as the Ro-Phon corpus.

In line with the first research question, our data show that creaky phonation in the context of filler particles is not restricted to only one type of filler (surfacing in vocalic, vocalic-nasal and in nasal outputs alike), nor is it the attitude of one gender, appearing in both male and female monologues. Furthermore, creaky fillers are not restricted to utterance-final position, emerging both in pre-pausal as well as post-pausal contexts.

Modal voice fillers displayed a flat f_0 contour, with the highest average value recorded at vocalic-nasal filler particles (197 Hz) and the lowest average value in nasal fillers (113 Hz), both of which occurred before a silent pause. In contrast, only the vocalic creaky filler showed a regular f_0, with irregular (vocalic-nasal) and highly irregular patterns in the context of pre-pausal nasal filler particles derived from male speech. Independent of filler structure and position in the utterance, the fundamental frequency constantly had a lower f_0 standard deviation in the case of modal phonation.

The data revealed that, across all filler particles, there are fewer f_0 periods per second in the creaky tokens compared to modal voice outputs. Furthermore, a higher degree of voice breaks demarcates creaky from non-creaky tokens, especially in male speech.

Our results support previous findings in the sense that jitter and shimmer are systematically higher in creaky fillers, with both highest values found in male speech (nasal filler particle). Likewise, HNR and CPP always presented lower values in creaky voicing, with both peaking in the vocalic-nasal filler token. As for spectral tilt measurements, the results are inconclusive. Consequently, in future studies we intend to use the formant-corrected amplitude difference between the first and second harmonics (H1*-H2*), with f_0 measured by means of the STRAIGHT algorithm in VoiceSauce (Shue et al. 2011). This approach is also motivated by the fact that the algorithm can accurately track low fundamental frequency values, characteristic of non-modal phonation.

With regard to the temporal domain, as a consequence of a small-scale analysis, results varied. In terms of voice quality, vocalic filler particles had an overall longer duration in modal phonation, vocalic-nasal outputs spanned across a similar temporal interval, whereas nasal fillers were longer in creaky phonation. As for the average frequencies of the first formant, results revealed that they were higher in creaky phonation.

Pinpointing the exact type of creaky voice proves to be rather challenging, especially when working on connected speech data characterised by abrupt shifts of the fundamental frequency (see section 2.2, figures 1–4). The complexity of the procedure is elevated since it takes only one feature, either a low f_0, an irregular f_0 or a constriction of the glottis for the voice to be perceived as creaky (feature previously observed by Davidson 2019, Keating et al. 2023, among others). In this regard, based on the classification proposed by Keating et al. (2015) (see section 2.2), Garellek (2019) provides a visual summary of different creak sub-categories, alongside the acoustic differences between modal, breathy and creaky phonation.

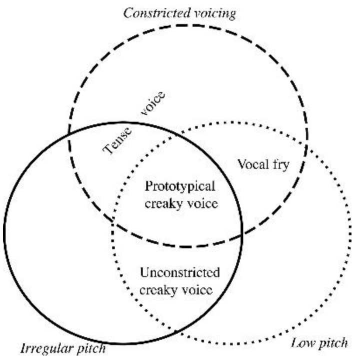

Figure 11
Creaky voice represented in a three-dimensional space.
Image taken from Garellek (2019: 81)

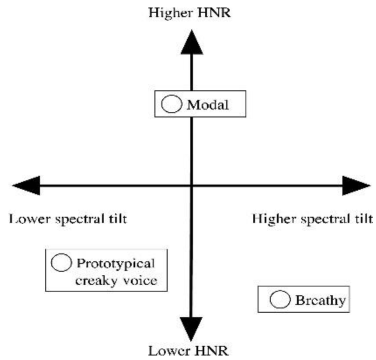

Figure 12
Acoustic differences among modal, breathy and (prototypical) creaky voice.
Image taken from Garellek (2019: 90)

Against this background, we can infer that the vocalic creaky filler from our study illustrates the qualities of vocal fry, having a low but regular f0 (unlike prototypical creaky which is defined as having an irregular fundamental frequency), the vocalic-nasal filler, showcasing a high H1-H2, resembles a nonconstricted creaky phonation, while the nasal filler particle denotes an aperiodic creak.

It should also be mentioned that all filler particles examined here are either completely modal or fully creaky outputs. However, this is not always the case, especially in spontaneous speech where fillers not only transition from modal to creaky phonation within the same token, but also overlap with other disfluency phenomena, forming complex clusters. Henceforth, future studies will take into account sequences of modal-creaky or creaky-modal filler particles and further explore how these outputs interact with different nonverbal vocalisations (Trouvain 2014), such as breathing noises, clicks (interpreted as paralinguistic signals) and laughter. The addressed monologues from the Ro-Phon corpus often exhibit all of the previously mentioned features, as can be seen in the following image:

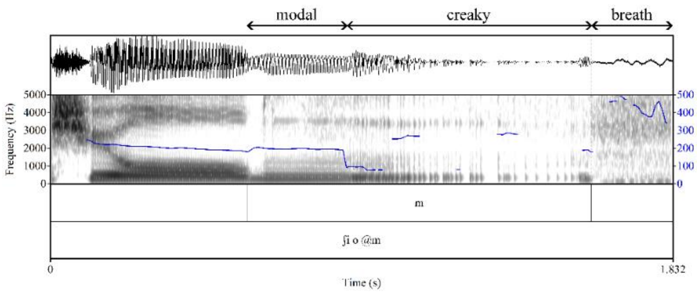

Figure 13
Filler particle with voice quality changing over time (Ro-Phon, S_5)

Figure 13 illustrates a nasal filler particle beginning in modal phonation and then transitioning into creaky voice, followed by breath intake.

The second research question addressed in our study relates to whether filler particles fulfil different pragmatic functions depending on voice quality. Before providing potential answers to this question, we need to acknowledge the limitations of the current analysis regarding sample size and data overview. Based on our preliminary evaluation of the monologues gathered in the Ro-Phon corpus, we noticed that creaky fillers are not exclusively used for conveying an attitude of uncertainty/indecisiveness, as seen with the Swedish filler particle "EH" (Horne 2009). While the discursive functions of these filler particles in terms of voice quality is still uncertain in the current stage of our research, it is worth noting that both phonation types are employed in male and female speech.

The distribution, frequency, and patterns of filler particles in Romanian speech remains to be explored in upcoming studies. In turn, this future work will help us gain a better understanding of the relationship between the form and the function of fillers within the discourse. Data derived from less-documented European languages, such as Romanian, can significantly contribute to broadening our understanding on the typology and functions of filler particles across different types of voice phonation in a cross-linguistic context.

Acknowledgments
The Ro-Phon corpus was developed during the Roc-lingv project supported by a grant of the Romanian Ministry of Education and Research, CNCS-UEFISCDI, project number PN-III-P1-1.1-PD-2019-1029, within PNCDI III. I would like to thank my colleague, Bianca Alecu (ORCID ID 0009-0008-2192-6730), for her valuable insights during our discussions about the different pragmatic functions filler particles have in the discourse. I am also grateful for the helpful comments given by the anonymous reviewers. Any remaining errors are our own.

REFERENCES

Barbu, A. M., E. Irimia, C. Mîrzea Vasile, V. Păiș, 2023, “Designing the LECOR Learner Corpus for Romanian”, in: G. Angelova, M. Kunilovskaya, R. Ritov (eds), *Deep Learning for natural Language Processing Methods and Applications* (*Proceeding of the 14th Conference Recent Advances in Natural Language Processing* (RANLP 2023), Varna), 143–152.

Belz, M., 2017, “Glottal filled pauses in German”, *Proceedings of the Eighth International Workshop on Disfluency in Spontaneous Speech* (DiSS 2017), Stockholm, Sweden, 5–8.

Belz, M., 2023, “Defining Filler Particles: A Phonetic Account of the Terminology, Form, and Grammatical Classification of “Filled Pauses””, in: J. Trouvain, B. Möbius (eds), *Special Issue Pauses in Speech*, *Languages*, 8, 57, 27 p.

Davidson, L., 2019, “The effects of pitch, gender, and prosodic context on the identification of creaky voice”, *Phonetica*, 76, 4, 235–262.

Davidson, L., 2021, “The versatility of creaky phonation: Segmental, prosodic, and sociolinguistic uses in the world's languages”, *Wiley Interdisciplinary Reviews: Cognitive Science*, 12, 3, e1547 (18 p.).

Garellek, M., 2019, “The phonetics of voice”, in: W, F. Katz, P. F. Assmann (eds), *The Routledge Handbook of Phonetics*, New York, Routledge, 2019, 75-106.

Gobl, C., A. NÍ Chasaide, 2010, “Voice source variation and its communicative function”, in: W.J. Hardcastle, J. Laver, F.E. Gibbon (eds), *The Handbook of Phonetic Sciences*, 2nd edition, Oxford, Blackwell, 378–423.

Gordon, M., P. Ladefoged, 2001, “Phonation types: A cross-linguistic overview”, *Journal of Phonetics*, 29, 4, 383–406.

Guimarães, I., 2007, *A Ciência e a arte da voz humana, Alcabideche*, Escola Superior de Saúde de Alcoitão.

Horne, M, 2006, "The filler EH in Swedish", *Proceedings of Fonetik* 2006, Working Papers, Department of Linguistics and Phonetics, Lund University, 52, 65–68.

Horne, M., 2009, "Creaky fillers and speaker attitude: data from Swedish", in D. Barth-Weingarten, N. Dehé, A. Wichmann (eds), *Where prosody meets pragmatics*, 8, Leiden, Brill, 277–288.

Hough, J., L. de Ruiter, S. Betz, D. Schlangen, 2015, "Disfluency and laughter annotation in a light-weight dialogue mark-up protocol", *Proceedings of the Seventh International Workshop on Disfluency in Spontaneous Speech* (DiSS 2015), Edinburgh, UK, 49–52.

Keating, P. A., M. Garellek, J. Kreiman, J., 2015, "Acoustic properties of different kinds of creaky voice", *Proceedings of The 18th International Congress of Phonetic Sciences* (ICPhS 2015), Glasgow Scotland, UK, 5 p.

Keating, P., M. Garellek, J. Kreiman, Y. Chai, 2023, "Acoustic Properties of Subtypes of Creaky Voice", poster presentation, *Acoustic Society of America*, 2 p.

Kirk, P. L., J. Ladefoged, P. Ladefoged, 1993, "Quantifying acoustic properties of modal, breathy and creaky vowels in Jalapa Mazatec", in: A. Mattina, T. Montler (eds), *American Indian linguistics and ethnography in honor of Laurence C. Thompson*, Missoula, MT, University of Montana Press, 435–450.

Ladefoged, P., I. Maddieson, 1996, *The sounds of the world's languages*, Cambridge, MA, Blackwells.

Laver, J., 1980, *The Phonetic Description of Voice Quality*, Cambridge, Cambridge University Press.

Ludusan, B., H. Mattias, W. Marcin, 2023, "Exploring the role of formant frequencies in the classification of phonation type", *Proceedings of the 20th International Congress of Phonetic Sciences* (ICPhS 2023), Prague, Czech Republic, 1750–1754.

Mîrzea Vasile, C., 2020, "The LECOR Project. A Presentation", in: V. Barbu Mititelu, E. Irimia, D. Tufiș, D. Cristea (eds), *Proceedings of the 15th International Conference "Language Resources and Tools for Natural Language Processing"*, Iași, Editura Universității A. I. Cuza, 37–48.

Munro, P., F. Lopez, 1999, *Di'csyonaary X:tèe'n Dìi'zh Sah Sann Lu'uc*, Los Angeles, UCLA Chicano Studies Research Center.

Niculescu, O., 2021, "Developing linguistic resources for Romanian written and spoken language", *Proceedings of the 16th International Conference* „Linguistic Resources and Tools for Natural Language Processing", Iași, Editura Universității „Alexandru Ioan Cuza", 21–36.

Niculescu, O., 2023, "Acoustic correlates of filler particles in Romanian connected speech", *Philologica Jassyensia*, in evaluation.

Schegloff, E. A., 2010. "Some Other 'Uh(m)'s", *Discourse Processes*, 47, 2, 130–174.

Shue, Y.-L., P. Keating, C. Vicenik, K. Yu, 2011, "VoiceSauce: A program for voice analysis", *Proceedings of The 17th International Congress of Phonetic Sciences* (ICPhS 2011), Hong Kong, China, 1846-1849.

Slifka, J., 2006, "Some Physiological Correlates to Regular and Irregular Phonation at the End of an Utterance", *Journal of Voice*, 20, 2, 171–186.

Teixeira, J. P., C. Oliveira, C. Lopes, 2013, "Vocal Acoustic Analysis – Jitter, Shimmer and HNR Parameters", *Procedia Technology*, 9, 1112–1122.

Teixeira, J. P., P. O. Fernandes, 2014, Jitter, "Shimmer and HNR classification within gender, tones and vowels in healthy voices", *Procedia Technology*, 16, 1228–1237.

Trouvain, J., 2014, "Laughing, breathing, clicking – The prosody of nonverbal vocalisations", *Proceedings of Speech Prosody*, 598–602.

Tylečková, L., R. Skarnitzl, 2019, "The mapping of voice parameters in connected speech of healthy Common Czech male speakers", *Akusticke listy*, 25, 1–2, 10–18.

Watts, C. R., S. N. Awan, 2011, "Use of Spectral/Cepstral Analyses for Differentiating Normal From Hypofunctional Voices in Sustained Vowel and Continuous Speech Contexts", *Journal of Speech, Language, and Hearing Research*, 54, 1525–1537.