forum acusticum 2026
8-12 September · Graz, Austria

TEMPORAL STRUCTURE OF CREAK IN PATHOLOGICAL READ SPEECH

Anna Viehhauser[1], Barbara Schuppler[1], Philipp Aichinger[2,3]

[1]*Signal Processing and Speech Communication Laboratory, Graz University of Technology, Austria*
[2]*Speech and Hearing Science Lab, Division of Phoniatrics and Speech Language Therapy, Department of Otolaryngology, Head and Neck Surgery, Medical University of Vienna, Austria*
[3]*Comprehensive Center for AI in Medicine, Medical University of Vienna, Austria*

This paper is a revised version of the authors' submitted manuscript that was peer-reviewed by at least two anonymous reviewers.

Abstract

Creak is a voice quality frequently occurring in healthy speech, sometimes carrying linguistic or paralinguistic information. Due to the lack of precision in voice control, pathological speakers may produce creak independently from its usual functions. This paper investigates the similarities and differences in temporal characteristics of creak occurences using recordings of German read speech. Based on automatically annotated and manually corrected creak intervals, our results show that pathological speakers creak longer and speak slower than healthy speakers, while the creak rate and the gap between two subsequent creaks remain similar. Analysis shows a loss of correlation between articulation rate and mean creak length as well as articulation rate and creak rate in pathological speech. Furthermore, we find that creak occurs at phrase and sentence boundaries in both groups. The pathological dataset shows a less distinct and more dispersed structure, which suggests reduced targeted voice control and more irregular creak onsets. The study provides new insights into the temporal characteristics of creak in pathological speech and how they reflect impaired voice control.

Keywords: *creak, pathological speech, voice quality, creaky voice, read speech*

1. Introduction

Voice quality plays an important role in speech analysis as it carries not only linguistic content but also information about the speaker's emotional condition [1] and health [2]. One particular voice quality is called creak and it is essential for phonological analyses and the investigation of turn-taking [3]. Keating et al [4] define the key properties of creak as a low vibration frequency of the vocal fold, an irregular fundamental frequency and a constricted glottis, which consists of a small peak glottal opening, a long closed phase and low glottal airflow. Since not all of these properties are always present, they propose to distinguish between the following subtypes: vocal fry, multiply pulsed voice, aperiodic voice, nonconstricted creak, and tense/pressed voice [4]. Similarly, Ashby [5] defines creak as a type of phonation where the vocal folds vibrate irregularly with a low vibration frequency and low airflow through the glottis. Usually, the fundamental frequency is below the range of the modal voice [6]. Auditory creak is perceived as repeating impulses with a rough quality [2] and is often described as a stick run along a railing [7]. Creak, as a symptom of pathology, is a subtype of perceived roughness [8]. For our work, we define creak as a voice quality with low vibration frequency of the vocal folds and larger amounts of jitter and shimmer than in modal speech. Jitter and shimmer are fast random fluctuations of cycle length and peak amplitude, respectively.

Creak is a normal variation in healthy speech and often occurs at phrase and sentence boundaries [9, 10] or when the vocal folds are very relaxed [2]. In pathological speech, creak can appear more prominently and persistently and may occur independently from its usual functions. This may happen due to a lack of targeted voice control, for example, less control of glottal closure or the vibration of the vocal folds. In general, pathological speech may have an increased perturbation (i.e., jitter and shimmer) and elevated noise components. The aim of this work is to increase our understanding of the differences in healthy and pathological speech with a specific focus on the temporal variation in creak occurences.

The present study addresses the variation of temporal voice quality focusing on automatically annotated

Corresponding author: *b.schuppler@tugraz.at*

and manually corrected creak intervals using *creapy* [3] and *Praat* [11]. Using recordings of German read speech, we analyze the features mean creak length, creak rate and how they affect the articulation rate, as well as the mean creak gap reflecting the time between two subsequent instances of creak. Since all speakers read the same text, we examine similarities and differences of creak occurences on a sentence level in healthy and pathological speech. With this approach, we aim at increasing the understanding of creak behavior in pathological speech. Discovered differences in temporal voice quality may support and supplement acoustic and perceptual assessments of disorders.

2. Materials and Methods

The dataset used for this project consists of 59 audio files from 36 (20 female and 16 male) pathological and 23 (12 female and 11 male) healthy speakers. The pathological speakers have different kinds of voice disorders, including edema, paresis, and cysts. Each of the speakers reads the German standard text "Der Nordwind und die Sonne", which consists of six sentences plus the title. The audio files of the pathological speakers and 14 of the healthy speakers were recorded at the Medical University of Vienna [12]. The remaining nine audio files of the healthy speakers were taken from the Graz Corpus of Read and Spontaneous Speech (GRASS), recorded at the Graz University of Technology [13]. Table 1 shows the number of speakers in the healthy and pathological dataset and the corresponding length of the audio files in seconds.

Table 1: Number of speakers in the healthy and pathological dataset (f corresponds to the number of female speakers, m to the number of male speakers) and the corresponding audio time in seconds.

	number of speakers	**audio time**
healthy	23 (12 f, 11 m)	1045.47 s
pathological	36 (20 f, 16 m)	1863.50 s
total	59 (32 f, 27 m)	2908.97 s

2.1. Annotation Process

The creak intervals are automatically annotated using *creapy* [3] and manually corrected in *Praat* [11]. This was done by listening to short intervals of the audio file and by performing a visual analysis of the waveform and the spectrogram by the first author. If the interval sounded creaky and the waveform and spectrogram had the properties described in Section 1, we labelled the interval as *c*. Each boundary is aligned with the nearest zero-crossing of the audio waveform.

2.2. Extraction of Temporal Features

The annotated creak intervals were extracted using the Python [14] packages *parselmouth* [15] and *textgridtools* [16]. We compute the following temporal features for each speaker:

- ***mean creak length (mcl):*** The mean creak length describes the average length of a creak interval for each speaker in seconds.
- ***mean creak gap (mcg):*** This feature describes the mean time gap between two creak intervals in seconds.
- ***articulation rate (ar):*** The articulation rate corresponds to the number of canonical phones per second (including pauses).
- ***creak rate (cr):*** The creak rate corresponds to the number of creaks divided by the spoken time (creaks per second).

3. Results and Discussion

3.1. How do healthy and pathological speakers differ regarding temporal features?

Table 2 compares the mean and standard deviation across the features *mean creak length* (mcl), *articulation rate* (ar), *creak rate* (cr) and *mean creak gap* (mcg) for the healthy and pathological dataset. The *mean creak gap* is computed only for speakers who creak more than once (18 healthy and 33 pathological speakers). We see that pathological speakers creak longer with a mean of about 97 ms, compared to about 72 ms in healthy speakers. They also speak slower with an *articulation rate* of about 13 phones per second (15 for healthy speakers). The creak rate is similar for both with 0.24 (healthy) and 0.28 (pathological) creaks per second. On average, the gap between two subsequent creaks is similar for both healthy and pathological speakers with 5.52 s (healthy) and 5.91 s (pathological).

Table 2: Population means and standard deviations of the *mean creak length* (mcl), *articulation rate* (ar), *creak rate* (cr) and *mean creak gap* (mcg) for the healthy and pathological data.

	mcl (ms)	**mcg** (s)	**ar** (phones/s)	**cr** (creak/s)
healthy	71.61 ±28.88	5.52 ±4.40	15.14 ±1.59	0.24 ±0.25
pathological	97.06 ±39.08	5.91 ±5.24	13.04 ±1.89	0.28 ±0.27

Figure 1 shows the mean (solid line) and standard deviation (shaded area) for the min-max normalized features *mean creak length* (mcl), *articulation rate* (ar), *creak rate* (cr) and *mean creak gap* (mcg), separately for the pathological (blue) and healthy (red) dataset. We can see that the normalized mean *mean creak gap* and mean *creak rate* are very similar for both datasets, unlike the mean *mean creak length* and the mean *articulation rate*, which show greater differences.

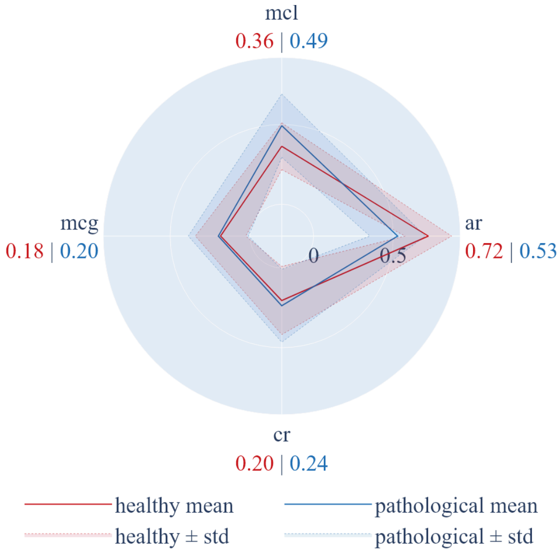

Figure 1: Comparison of the mean (solid line) and the standard deviation (shaded area) of the normalized features *mean creak length* (mcl), *articulation rate* (ar), *creak rate* (cr) and *mean creak gap* (mcg) for the pathological (blue) and healthy (red) dataset.

To further investigate how these features correlate, we take a look at the correlation matrices for both healthy and pathological speakers. Figure 2a and Figure 2b show correlation matrices for the healthy (red) and the pathological (blue) dataset with the features *mean creak length* (mcl), *articulation rate* (ar), *creak rate* (cr) and *mean creak gap* (mcg). For the healthy data, there is a moderate positive correlation between the *articulation rate* and the *mean creak length* (0.41) and a similar correlation between *articulation rate* and *creak rate* (0.37), suggesting that speakers who speak faster creak longer and more frequently. In contrast to this, the correlation between *articulation rate* and *mean creak length* as well as between *articulation rate* and *creak rate* drop to near zero, indicating a decoupling of the *articulation rate* and creak production. This may reflect reduced targeted voice control or irregularities in creak onsets. For both datasets, the *mean creak gap* and the *creak rate* achieve the highest correlation score with −0.69 for the healthy and −0.51 for the pathological dataset, indicating that speakers who creak more often also have shorter gaps in between. This correlation is statistically significant in both datasets, with a stronger significance in the healthy data ($p = 0.0015$) compared to the pathological data ($p = 0.0022$). In the healthy dataset, the *mean creak gap* and the *articulation rate* negatively correlate (-0.33) but for the pathological speakers there is a slight positive correlation with 0.26. This indicates that a lower *articulation rate* leads to a larger gap between two subsequent creak intervals in the healthy dataset, but in the pathological dataset speakers who speak slower also have smaller gaps between two creak intervals. Another difference can be observed for the *mean creak gap* and the *mean creak length*. There is almost no correlation for pathological speakers, but a moderate and statistically significant correlation for healthy speakers (-0.59). Overall, healthy and pathological speakers differ in the correlation of the *articulation rate* and the creak features (*mean creak length*, *creak rate*, and *mean creak gap*).

3.2. How does the articulation rate affect the creak length?

Based on the correlation matrices, where the difference between pathological and healthy speakers is large with respect to the correlation between *articulation rate* and *mean creak length*, we further investigate how they affect each other. To gain more insight on how the *articulation rate* does affect the creak length, Figure 3 shows the *mean creak length* over the *articulation rate* for pathological and healthy speakers. Each of the bubbles represents one speaker, the color corresponds to the health status (blue for pathological, red for healthy) and the size corresponds to the number of creaks produced by each speaker. The minimal number of creaks for both datasets is zero (see the four bubbles on the bottom of Figure 3). The maximal

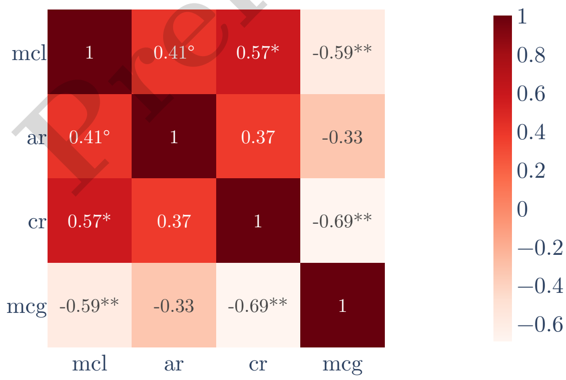

Figure 2a: healthy

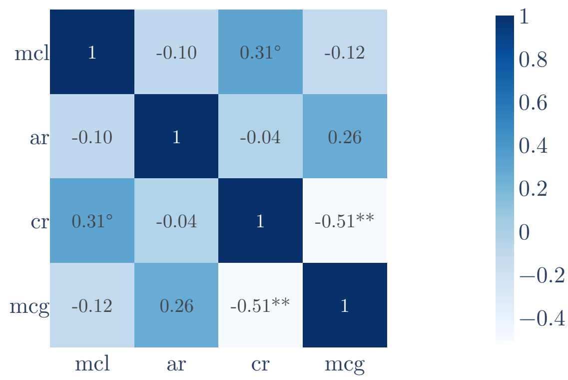

Figure 2b: pathological

Figure 2: Correlation matrix for the healthy and pathological dataset with the features *mean creak length* (mcl), *articulation rate* (ar), *creak rate* (cr) and *mean creak gap* (mcg). The stars indicate the p-value p, ** correspond to $p < 0.01$, * to $p < 0.05$, and ° to $p < 0.1$.

number of creak occurences in the healthy dataset is 25, and 46 in the pathological dataset.

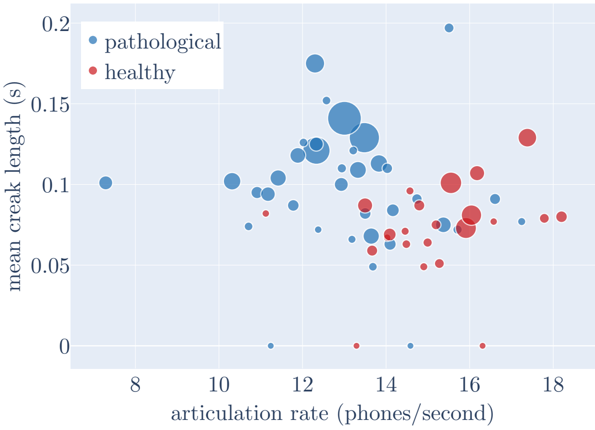

Figure 3: *Mean creak length* (in seconds) over *articulation rate* (in phones per second) for the healthy (red) and pathological (blue) dataset. Each bubble represents a speaker, the size corresponds to the number of creaks produced by each speaker.

As in Table 2 and Figure 1, we see that healthy speakers tend to speak faster than pathological speakers. On average, the pathological speakers creak longer and more often than healthy speakers. Additionally, the bubbles of the pathological dataset are more widely distributed than the healthy ones, an observation consistent with the results of the correlation of the *mean creak length* and the *articulation rate* in Figure 2a (0.41) and Figure 2b (-0.10). All in all, pathological speakers produce creaks that are longer, more frequent, and more variable than creaks produced by healthy speakers.

3.3. Creak occurences in a normalized timeline

To compare the temporal location of creak occurences in the course of the audio file over different healthy and pathological speakers, we normalize the audio duration. By doing so, we assume that the articulation rate is approximately equal across speakers. Since we only account for the presence of a creak and not the length, the center of each creak interval was used as time point for creak occurence. Since two pathological speakers (one female, one male) did not read the title of the reading text, they are excluded from this figure. The result is shown in the histograms of Figure 4. The top plot in blue shows the result for the pathological speakers, the plot on the bottom in red corresponds to the healthy data. The dotted part corresponds to female speakers, the striped area to male speakers. The lines correspond to phrase (thin) and sentence boundaries (thick). The timesteps of the boundaries are extracted from the TextGrids, normalized by the length of the audio and averaged over the speakers.
First, there are clear peaks visible in both groups, so creak occurences are not distributed uniformly over time, which may suggest that creak occurs more frequently at phrase boundaries [9, 17]. In the bottom plot, we can discern the structure of the six read sentences, by assuming a peak is located at the end of a sentence. The structure for the pathological dataset is very similar, but with a slight offset. Additionally, the peaks are not as distinct as for the healthy dataset, indicating less controlled and more irregular creak onsets. At area A (at 0.6-0.65 on the time axis) there

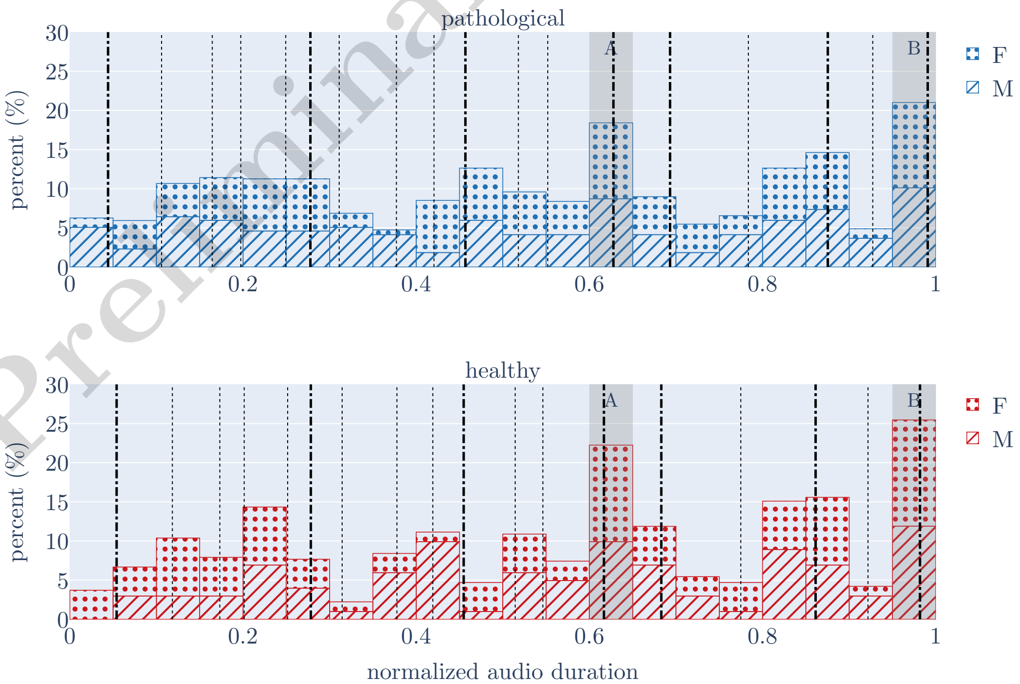

Figure 4: Histogram of creak occurence (stacked data of female (F) and male (M) speakers) over normalized audio duration for pathological speakers on top (blue) and healthy speakers on the bottom (red). Thin lines correspond to phrase boundaries, thick lines to sentence boundaries.

is a clear peak in both groups indicating a phrase or sentence boundary. In fact, this peak corresponds to the phrase "... desto fester hüllte sich der Wanderer in seinen Mantel ein.". The high peak at area B (at 0.95-1 on the time axis) corresponds to the end of the last sentence "... dass die Sonne von ihnen beiden der Stärkere war.". This suggests that creak occurs more frequently at the end of a phrase.

Regarding the differences between male and female speakers, healthy male speakers tend to cluster creak at certain points rather than spreading it more evenly over all utterances like pathological male speakers. Most of the time, the amount of creak is gender balanced, but at 0.4-0.45 the pathological female speakers creak more often than the pathological male speakers. At this time point in the healthy dataset, the opposite is true: most of the occurring creaks are produced by male speakers. The area around 0.4 corresponds to the text "... der den Wanderer zwingen würde, seinen Mantel abzunehmen." At area A and B, the creak occurence is gender balanced in both groups. Overall, creak occurences are concentrated at phrase and sentence boundaries in both groups, but pathological speakers show less distinct, more irregular peaks.

These findings should be interpreted in light of two limitations. First, all annotations were performed by a single annotator, making it impossible to assess inter-annotator agreement and potentially introducing annotator-specific bias. Second, the healthy dataset comprises two different sources, which may introduce corpus-specific variability independent of speaker health and could therefore contribute to differences in temporal features and their correlations.

4. Conclusion

In conclusion, creak is present in healthy and pathological groups but shows distinct temporal patterns. On average, pathological speakers creak longer and have a slower articulation rate. The gap between two subsequent creaks and the creak rate is similar, indicating that not the frequency but the temporal distribution and duration differentiate pathological from healthy speech. In the correlation analysis of the healthy dataset, the articulation rate correlates moderately with the mean creak length and the creak rate, whereas in the pathological dataset this correlation vanishes. If we normalize the timeline, we see that both groups show high creak occurences at structurally relevant points such as phrase and sentence boundaries. However, pathological speakers show less defined peaks of creak occurences and a greater variation in the timing of creak, indicating a reduced targeted voice control and a more irregular production of creak. Whereas, pathological female and male speakers show overall similar patterns of creak occurences, pathological female speakers show more temporal clustering of creak occurences.

The results support the assumption that the temporal structure of creak can serve as a marker to distinguish healthy from pathological speech and may supplement diagnosis and therapeutic assessments. Overall, this study deepens the understanding of temporal voice qualities and show that temporal creak features complement acoustic and perceptual assessments of speech disorders.

5. Acknowledgements

This work was supported by the Austrian Science Fund (FWF): KLI 722-B30.

References

[1] C. Gobl and A. N. Chasaide, "The role of voice quality in communicating emotion, mood and attitude," *Speech Communication*, vol. 40, no. 1–2, pp. 189–212, 2003, doi: https://doi.org/10.1016/S0167-6393(02)00082-1.

[2] C. T. Ishi, K.-I. Sakakibara, H. Ishiguro, and N. Hagita, "A Method for Automatic Detection of Vocal Fry," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 1, pp. 47–56, Jan. 2008, doi: 10.1109/TASL.2007.910791.

[3] M. Paierl, T. Röck, S. Wepner, A. Kelterer, and B. Schuppler, "Creapy: A Python-based tool for the detection of creak in conversational speech," 2023, [Online]. Available: https://guarant.cz/icphs2023/59.pdf

[4] P. Keating, M. Garellek, J. Kreiman, and others, "Acoustic properties of different kinds of creaky voice.," vol. 1, pp. 2–7, 2015.

[5] M. Ashby, *Introducing Phonetic Science*, Fifth printing. in Cambridge Introductions to Language and Linguistics. Cambridge [u.a.]: Cambridge University Press, 2010.

[6] J. H. Esling, S. R. Moisik, A. Benner, and L. Crevier-Buchman, *Voice Quality: The Laryngeal Articulator Model*. Cambridge University Press, 2019. doi: https://doi.org/10.1017/9781108696555.

[7] V. Devaraj, I. Roesner, F. Wendt, J. Schoentgen, and P. Aichinger, "Auditory Perception of Impulsiveness and Tonality in Vocal Fry," *Applied Sciences*, vol. 13, no. 7, p. 4186, Mar. 2023, doi: 10.3390/app13074186.

[8] P. H. Dejonckere *et al.*, "A basic protocol for functional assessment of voice pathology, especially for investigating the efficacy of (phonosurgical) treatments and evaluating new assessment techniques," *European Archives of Oto-Rhino-Laryngology*, vol. 258, no. 2, pp. 77–82, Feb. 2001, doi: https://doi.org/10.1007/s004050000299.

[9] J. Kreiman, "Perception of sentence and paragraph boundaries in natural conversation," *Journal of Phonetics*, vol. 10, no. 2, pp. 163–175, 1982, doi: https://doi.org/10.1016/S0095-4470(19)30955-6.

[10] L. Redi and S. Shattuck-Hufnagel, "Variation in the realization of glottalization in normal

speakers," *Journal of Phonetics*, vol. 29, no. 4, pp. 407–429, 2001, doi: https://doi.org/10.1006/jpho.2001.0145.

[11] P. Boersma and D. Weenink, "Praat: doing phonetics by computer." [Online]. Available: https://www.praat.org/

[12] P. Aichinger and J. Schoentgen, "Detection of Diplophonation in Audio Recordings of German Standard Text Readings," *Journal of Voice*, vol. 33, no. 6, p. 949.e1–949.e10, Nov. 2019, doi: 10.1016/j.jvoice.2018.06.009.

[13] B. Schuppler, M. Hagmüller, J. A. Morales-Cordovilla, and H. Pessentheiner, "GRASS: the Graz corpus of Read And Spontaneous Speech.," in *LREC*, 2014, pp. 1465–1470.

[14] G. Van Rossum and F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009.

[15] Y. Jadoul, B. Thompson, and B. de Boer, "Introducing Parselmouth: A Python interface to Praat," *Journal of Phonetics*, vol. 71, pp. 1–15, 2018, doi: https://doi.org/10.1016/j.wocn.2018.07.001.

[16] H. Buschmeier and M. Wlodarczak, "TextGridTools: A TextGrid processing and analysis toolkit for Python," in *Tagungsband der 24. Konferenz zur elektronischen sprachsignalverarbeitung (ESSV 2013)*, 2013.

[17] M. Garellek, "Perception of glottalization and phrase-final creak," *The Journal of the Acoustical Society of America*, vol. 137, no. 2, pp. 822–831, 2015, doi: https://doi.org/10.1121/1.4906155.