INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Inter-speech Clicks in an Interspeech Keynote

Jürgen Trouvain[1], *Zofia Malisz*[1,2]

[1] Computational Linguistics and Phonetics, Saarland University, Saarbrücken, Germany
[2] Department of Speech, Music and Hearing, KTH Stockholm, Sweden
trouvain@coli.uni-saarland.de, malisz@kth.se

Abstract

Clicks are usually described as phoneme realisations in some African languages or as paralinguistic vocalisations, e.g. to signal disapproval or as sound imitation. A more recent discovery is that clicks are, presumably unintentionally, used as discourse markers indexing a new sequence in a conversation or before a word search. In this single-case study, we investigated more than 300 apical clicks of an experienced speaker during a keynote address at an Interspeech conference. The produced clicks occurred only in inter-speech intervals and were often combined with either hesitation particles like "uhm" or audible inhalation. Our observations suggest a link between click production and ingressive airflow as well as indicate that clicks are used as hesitation markers. The rather high frequency of clicks in the analysed sections from the 1-hour-talk shows that in larger discourse, the time between articulatory phases consists of more than silence, audible inhalation and typical hesitation particles. The rather large variation in the intensity and duration and particularly the number of bursts of the observed clicks indicates that this prosodic discourse marker seems to be a rather acoustically inconsistent phonetic category.

Index Terms: pauses, clicks, discourse markers, speech style

1. Introduction

Recent studies on English spontaneous speech data [1-4] showed that a great majority of investigated speakers produced click sounds. The click rates however, varied greatly between speakers and exhibited substantial acoustic variation. Our study investigates clicks in the speech of a *single* person who produced a *large* number of clicks during a conference keynote address given in English.

1.1. Different usage of clicks

Clicks are well known to have phonemic status in various African languages [5-9]. They were also described as paralinguistic vocalisations [10] that signal disapproval, often spelled 'tsk-tsk' in English, and dissatisfaction [11, 12] or that imitate environmental sounds, e.g. horse-shoes on cobblestones. Further uses of clicking include articulatory techniques in beatboxing [13] and compensatory articulation for plosives and affricates in pathological speech [14]. In addition, tongue-clicking is used in echolocation, often by the blind, where trained persons deliberately produce click sounds as a "sight-by-sound-strategy" in order to detect obstacles [15].

1.2. Discourse clicks

In the aforementioned uses of clicks one can assume that the speakers are *aware* that they produce these sounds. In contrast, clicks are most probably *unconsciously* produced by many speakers in inter-speech pauses in spontaneous conversation. These inter-speech clicks are used as discourse markers with the function of either indexing a new sequence or signaling formulation difficulties, e.g. during word search. The evidence for discourse clicks is recent and was reported for English [1-4, 11], German [16] and Wolof [17].

Recordings of the same speaker producing a longer stretch of speech in the same, or comparable, speaking style and situation provide a good opportunity to investigate discourse clicks with regards to their distribution, acoustics and timing. Particularly the vicinity to breath markers and filled pauses is of interest [1, 2].

1.3. Phonetic characteristics of clicks

Clicks are often perceptually salient sounds. Phonemic clicks are produced by a closure in the anterior part of the vocal tract, e.g. the tip of the tongue at the alveolar ridge, and a closure of the back of the tongue and the velum. Using ingressive airstream, the release of the anterior closure generates a burst-like sound [5]. A later closure release at the velar place of articulation has led to the term *velaric airstream* (see also the debate on velaric vs. lingual airstream mechanisms in phonemic clicks in [8, 9]).

There is no good reason to believe that for discourse clicks speakers produce a closure at the velum. The anterior closure in the dental-alveolar region can be explained with a gesture for speech preparation [18]. Nevertheless, Wright [2] auditorily distinguishes discourse clicks between those with- and without velaric airstream.

One big difference between phonemic and discourse clicks is that the former occur as consonants in connected speech whereas the latter as isolated sounds in pauses. Consonant inventories of click languages usually show clicks at various places of articulation (bilabial, dental, and alveolar) and with different manners of release, e.g. lateral release. Participation of the voice, addition of aspiration, coupling of the nasal cavity, and voice quality such as breathy voice [9] were also observed. In consequence, click languages probably possess a substantially larger click sound inventory than the International Phonetic Alphabet is able to provide symbols for.

Places of articulation can also differ among discourse clicks as Wright [2] attested for her English data which included bilabial clicks. Interestingly, in some conversational corpora *lip smacks* are annotated - as a rather frequent non-verbal vocalisation [19]. In other corpora lip smacks are completely lacking [19] or they are annotated *along with* tongue clicks [20]. The perceptual and/or acoustic description of lip smacks and bilabial clicks is missing. It is also unknown whether discourse clicks involve bilabial clicks. Audiovisual and articulatory data would probably provide reliable answers to these questions.

Various researchers note that it is sometimes hard to clearly determine a click's place of closure or the type of airstream used [2-4, 21]. Ogden [4] mentions *percussives* as

http://dx.doi.org/10.21437/Interspeech.2016-1064

another category of click-like sounds in discourse. Ball [22] describes percussives as a part of a click and suggests the term 'click-cluck' for the element sequence in such a sound.

Finally, a further instantiation of non-phonemic clicks should be mentioned. In connected speech *weak clicks* can occur as a coarticulatory by-product when consonants with a closure at the alveolar ridge are followed by a consonant with a velar closing gesture. These low-intensity clicks were described as epiphenomena in French [23], English [24] and German [25, 26].

1.4. Hypotheses

It was the subjective impression of the first author when he listened to the presently investigated Interspeech keynote that the speaker produced a high number of clicks. Thus the hypothesis is that the click rate in this speech, i.e. number of clicks per minute, is higher than reported in other studies. For example, Gold et al. [3] describe individual click rate variability in 100 speakers and finds an average of four and the maximum at eleven clicks per minute.

Wright [2] observed that many discourse clicks were produced adjacent to inhalation noises in her data. Elsewhere we [27] suggest a connection between clicks and inhalation that combines i) ingressive airstream by inhalation in breath pauses, ii) apical "speech-ready" posture of the tongue, iii) a sudden vertical movement of the larynx downwards, and iv) a wide opening of the glottis. Thus, a stronger in-breath at inter-speech pauses would favour a click. The higher the prosodic break (expressed as a pause) in the prosodic hierarchy, e.g. between two paragraph-like discourse sections, the more likely it is that the pause is marked with a stronger inhalation. Therefore, we expect that a substantial amount of clicks co-occur with audible breathing noises.

Another important observation of Wright [2] is that clicks occur in the vicinity of fillers, i.e. hesitation particles in "filled pauses" such as 'uh' or 'uhm' in English. Fillers usually index formulation difficulties which may ocurr due to lexical search, syntactic construction or to signal new information. Consequently, we also expect proximity to fillers.

In slide presentation, a new sequence is visibly marked by changing from one slide to another (mostly by a mouse click). For this reason we hypothesise that most of the presentation slide changes will be accompanied with tongue clicks.

Regarding the acoustic characteristics there are no concrete hypotheses, e.g. differences compared to realisations of phoneme clicks. We can also assume a high variability in the acoustic characteristics within individuals since there is no phonemic contrast to other clicks. Therefore we concentrate on the exploration of the variability of parameters like duration, intensity, and centre of gravity (COG). This includes also the number of burst elements (see Figure 1) since as Wright [2] observed, click events are sometimes doubly articulated.

2. Method

2.1. Data

We investigated the invited keynote lecture at the Interspeech conference by Anne Cutler from the year 2014 [28]. As an accomplished scientist, the speaker is very experienced with respect to giving talks, including keynotes, to large audiences. A lecture such as a keynote speech can also be subsumed under *talk-in-interaction* because listeners are immediately addressed and, in this case, at some places, also directly asked a question by the speaker. The lecture was audio-/video-taped by the company Superlectures as commissioned by the local conference organisers. The material was made available to the authors on request and the consent of the recorded speaker to use the material for this study was obtained. The material included the video in mp4 format, the audio files (from the speaker's fixed microphone) in mp3 format, the slide show (24 slides in total), and automated subtitles. The duration of the entire recording was 58 minutes.

There were 13 various phases of the talk selected for further analysis (see Figure 2): words of thanks for getting an award (phase 1), details about affiliations (phase 2), a general intro to the upcoming talk (phase 3). Other sections are taken randomly from the body of the talk (4-10). The phase duration is about 3 minutes, resulting in 38 minutes total time analysis.

2.2. Labelling

The inter-speech pauses in the aforementioned 13 phases were segmented in Praat. The following categories of speech and non-speech were labelled as they occurred either in the pause or in the immediately preceding or following signal:

- *speech*,
- *click events*,
- *individual click bursts*,
- *fillers* ("uh", "uhm", lengthened "and", "that" "but", and combinations of "and/but/that" and "uh/uhm"),
- *silence*,
- *audible inbreaths*.

Please note that *audible inbreath* does not necessarily include all cases of breathing because of silent breathing [27]. Examples of clicks are illustrated in Figure 1.

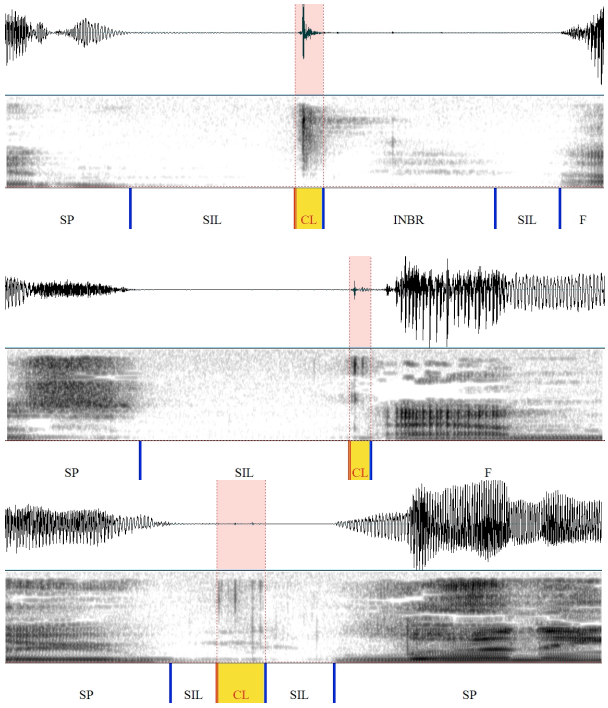

Fig. 1: *Waveforms and spectrograms (0-10 kHz) of 1 second long sections with clicks as examples. Top: single burst click (CL) followed by inbreath (INBR), silence (SIL) and a filler (F). Middle: double burst click preceded by silence and speech (SP). Bottom: quadruple burst click sandwiched between silence and speech.*

The clicks in the data were segmented by the authors. To establish whether the annotators recognise the same acoustic events as clicks, we compared their respective annotations of the same 3 min. long, randomly selected phase. We calculated event agreement [29] by counting the number of click events that overlapped between the annotators (n=96), multiplying it by two and dividing by the sum of all intervals from both annotators (n=111) in this phase. The resulting agreement equals 86%, indicating a high coherence between the annotators.

We cross checked the click events by examining the video recording to make sure that no manual mouse clicks were annotated as click events. We assumed no manual click was being used if the speaker was looking directly at the audience and/or her right hand was visible (usually in a co-speech gesture) or if the click was evident from orofacial visual cues. One click was subsequently corrected as actually swallowing.

We were interested in a possible temporal alignment of oral clicks and presentation clicks used to move to the next presentation slide. The moment of manual clicking was annotated as determined by the moment of the slide change in the full video. This procedure leads only to a rough approximation of the moment of the slide change.

3. Results

3.1. Click rates

Click rate was quantified as clicks per minute (cpm) speaking time (including all pauses). This measure was applied to determine whether there are substantial differences between phases and to compare click rates with those reported in other studies. Figure 2 shows the click rates, the mean for this speaker is 9.2 clicks per minute (the total number of all clicks in all phases is 323). It can be seen that in some phases, e.g. introduction, click rate is much higher than in others. The click-pause ratio indicates how many of the pauses found in a given phase contained a click.

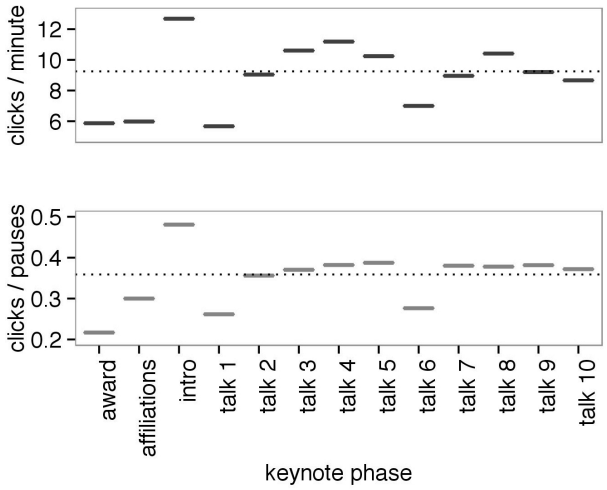

Fig. 2. *The click rate (clicks per minute, upper panel) and the number of click events relative to the number of pauses in a keynote phase (click to pause ratio, lower panel). The dotted lines represent the mean.*

3.2. Vicinity to audible inbreath noises and fillers

Audible inbreath noises before and/or after clicks could be observed in nearly half of the cases (46%), the majority of pauses with clicks did not contain inbreath noises (54%).

Forty three percent (n=139, Figure 3) of all inter-speech pauses containing clicks were followed (19%), preceded (14%) or abutted on both sides by fillers (10%).

Figure 4 illustrates the position of a click relative to the inter-speech pause it occurs in. The median values show that clicks are generally produced in the second half of the pause towards the onset of the following utterance. This late location is irrespective of whether a filler is involved in the click's immediate context or not.

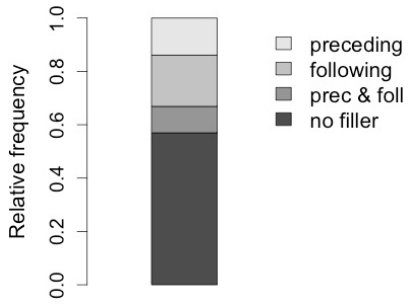

Fig. 3: *Number of inter-speech pauses in which a filler precedes and/or follows a click (or not).*

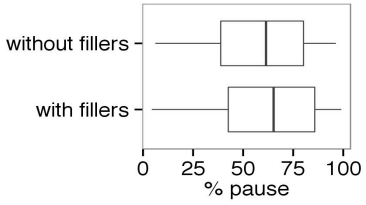

Fig. 4: *Time point of the click in an inter-speech pause either with or without a filler in its immediate context.*

3.3. Temporal alignment with presentation clicks

As illustrated in Table 1, 19 out of 23 slide changes were accompanied with a click 2 sec before and/or after the display of the change (as the presumed time point of presentation click). In total there were 12 clicks within a 2 sec-period *before* the slide change, and 11 clicks within the 2 sec-period *after* the slide change.

Table 1. *Temporal distance of tongue clicks before and after presentation clicks.*

time bin	before	after
0-1 sec	8	6
1-2 sec	4	5
2-3 sec	1	4
3-4 sec	0	3
>4 sec	10	4

3.4. Number of bursts

Table 2 illustrates that the majority of the investigated click events show more than one burst. Taken together, clicks with either one or two bursts made up 79% of the cases.

Table 2. *Amount of clicks according to the number of bursts.*

number of bursts	1	2	3	4	5
frequency	112	129	57	16	8
percentage	35%	40%	18%	5%	2%

3.5. Acoustic characteristics

Table 3 presents the duration summary of entire click events (single and multiple bursts treated as one event). Duration is quite variable (sd=25 msec.) exhibiting a wide range.

To analyse intensity and determine the centre of gravity (the "average frequency" of the spectrum) the strongest burst of each click event was examined. The strongest burst was selected on the basis of the waveform. The original signal was sampled at 22050Hz. The COG was computed via FFT using a 25 msec. Hamming window centred at the onset, midpoint and offset of the burst segment. Table 3 displays the values of maximum intensity, the rise time from the onset to maximum intensity (as in [8]) and COG averaged over the three burst phases.

Table 3. *Acoustic measures referring to the entire click event [ec] or to the strongest burst [sb].*

acoustic measure		min	mean	sd	max
duration (msec.)	[ec]	5	40	25	145
max. intensity (dB)	[sb]	31	51	9	73
rise time to max. intensity (dB/ms)	[sb]	2	8.2	5	25
COG (kHz)	[sb]	1.8	6.5	1.7	9.4

4. Discussion

The click rate clearly shows that the speaker ranges among the topmost "clickers" (upper 3%) if compared to the individual variability study by Gold et al. [3]. We also found differences in click rates along the phases of the talk: the highest rates in the introduction and in the first parts of the talk proper. In general, more than one third of the pauses in the analysed data contains a click, in one phase, even half of the pauses. We also found click rates lower than speaker average in some phases. After we completed the analysis, personal communication with the speaker revealed that she had rehearsed the award acceptance speech and that some studies she discussed in the keynote had been presented by her several times at other occasions.

Four out of ten pauses with clicks also occur in the immediate proximity of fillers. This strengthens the interpretation that clicks also act as markers of hesitations. Future work will show whether the general filler frequencies correlate with those of the clicks. If so, further evidence for clicks defined as indices of disfluent phases (among other parameters) would be provided.

The number of clicks co-occurring with an inbreath noise is substantial but lower than the figure given in Wright [2] (46% vs. 62%), who studied different discourse styles in different speakers and in several corpora. This result can be explained not only by individual variation but also the correlated effects of speech rate and speaking style.

As hypothesised, there is a close relationship between presentation clicks to change a slide and tongue clicks. Only four out of 23 slide changes were not accompanied by a tongue click within a 2 sec-period.

The examined acoustic profile of click events and their most prominent component bursts shows a high variability in duration, COG and intensity measures. The rather large variation in the intensity and duration and particularly the number of bursts indicates that clicks as a prosodic discourse marker seem to be a rather acoustically inconsistent phonetic category.

The COG values of the conversational clicks found in our speaker show energy concentrated in the higher part of the spectrum, similarly to values typically observed in pulmonic voiced fricatives [29]. The average value is higher (6.5kHz) than the average COG for the /t/-burst in English (4.5 – 5kHz) [31, 32]. The probable place of articulation of the clicks observed in our speaker is anterior. The high COG values would also point to a constriction in the alveolar to dental region, in accordance with the generalisation that front constrictions have higher spectral peaks.

5. Conclusions

This study has confirmed that the frequency of occurrence of tongue clicks in talk-in-interaction can be very high indeed. Though the investigated data does not involve a dialogue in its typical sense, it clearly contains features of interaction. As expected, this keynote shows many characteristics of spontaneous speech which is reflected by a substantial number of hesitations which are also marked with clicks.

This study also demonstrated that inter-speech pauses contain more information than just silence and fillers. Breathing noises and clicks also contribute to a large degree to the stretches of a talk which is not speech but nevertheless indispensable for discourse structuring and prosodic phrasing [27].

The perceptual salience typical for phonemic and paralinguistic clicks seems to be diminished in inter-speech clicks in discourse. Similarly, the great acoustic variation might not contribute to the reliable perceptibility of discourse clicks. Questions remain as to the perceptual relevance of click events in parsing discourse. There are many details which should be explored in future studies. Our immediate goal is to investigate the possible influence of surrounding phoneme articulations on the occurrence of clicks.

This study confirmed that inter-speech clicks can be a highly speaker-dependent feature making it relevant for various fields of automatic speaker characterisation. Studies in this direction would require a large number of speakers.

Other analyses such as the integration of visual parameters, the impact of the syntactic location (within or between sentences and "paragraphs"), and an improved scheme to interpret pragmatic and discourse functions are needed in different speaking styles. Also data from more languages would be useful to check whether discourse clicks have a universal character.

6. Acknowledgements

The authors would like to thank Anne Cutler for agreeing to use her speech samples and for her feedback on this paper; the company *Superlectures* for technical support with the material; four anonymous reviewers for their valuable comments. This research was supported by the ICT-TNG postdoctoral grant "Enhanced prominence modelling for the WikiSpeech open source speech synthesizer" to Zofia Malisz.

7. References

[1] M. Wright, "Clicks as markers of new sequences in English conversation", in *Proceedings of 16th ICPhS – International Congress of Phonetic Sciences, Saarbrücken*, pp. 1069–1072, 2007.

[2] M. Wright "On clicks in English talk-in-interaction", *Journal of the International Phonetic Association,* vol. 41, no. 2, pp. 207–229, 2011.

[3] E. Gold, P. French, and Ph. Harrison, "Clicking behaviour as a possible speaker discriminant in English", *Journal of the International Phonetic Association*, vol. 43, no. 3, pp. 339-349, 2013.

[4] R. Ogden, "Forms and functions of clicks in English conversation", *Journal of the International Phonetic Association*, vol. 43, no. 3, pp. 299-320, 2013.

[5] P. Ladefoged, *A Course in Phonetics*. 5th ed. Boston: Thomson/ Wadsworth, 2006.

[6] I. Maddieson, *Patterns of Sounds*. Cambridge: Cambridge University Press, 1984.

[7] I. Maddieson, "Presence of uncommon consonants", in M.S. Dryer and M. Haspelmath (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/ Accessed on 2014-01-14.), Chapter 19, 2013.

[8] A. Miller and S. Shah, "The acoustics of Mangetti Dune !Xung clicks", in *Proceedings of Interspeech, Brighton*, 2283-2286, 2009.

[9] A. Miller, "The representation of clicks", in M. van Oostendorp, C. Ewen, E. Hume, and K. Rice (eds). *Companion to Phonology*. Chichester: Wiley-Blackwell, pp. 416-439, 2011.

[10] D. Gil, "Para-linguistic usages of clicks" in M.S. Dryer and M. Haspelmath (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/ Accessed on 2014-01-14.), Chapter 142, 2013.

[11] N. Ward, "Non-lexical conversational sounds in American English", *Pragmatics and Cognition*, vol. 14, no. 1, pp. 113-184, 2006.

[12] E. Reber, *Affectivity in Interaction: Sound Objects in English*. Amsterdam/ Philadelphia: John Benjamins, 2012.

[13] M. Proctor, E. Bresch, D. Byrd, K. Nayak and S. Narayanan, "Paralinguistic mechanisms of production in human 'beatboxing': A real-time MRI study", *Journal of the Acoustical Society of America*, vol. 133, pp. 1043-1054, 2013.

[14] F. Gibbon, A. Lee, I. Yuen and L. Crampin, "Clicks produced as compensatory articulations in two adolescents with velocardiofacial syndrome", *The Cleft-Palate Craniofacial Journal*, vol. 45, no. 4, pp. 381-392, 2008.

[15] L. Wallmeier and L. Wiegrebe, " Self-motion facilitates echo-acoustic orientation in humans", *Royal Society Open Science*, Nov. 2014 [DOI: 10.1098/rsos.140185]

[16] J. Trouvain, "On clicks in German", in A. Leemann, M.-J. Kolly, S. Schmid and V. Dellwo (eds) *Trends in Phonetics and Phonology. Studies from German-speaking Europe*. Frankfurt/M. & Bern: Peter Lang, pp. 21-33, 2015.

[17] L.A. Grenoble, M. Martinović, and R. Baglini, "Verbal gestures in Wolof", in *Selected Proceedings of the 44th Annual Conference on African Linguistics*, Cascadilla Proceedings Project, Somerville, MA, pp. 110-121, 2015.

[18] J.M. Scobbie, S. Schaeffler and I. Mennen, "Audible aspects of speech preparation", in *Proceedings of 16th ICPhS – International Congress of Phonetic Sciences, Hong Kong*, pp. 1782-1785, 2011.

[19] J. Trouvain, and K.P. Truong, "Comparing non-verbal vocalisations in conversational speech corpora*",* in *Proceedings of the 4th International Workshop on Corpora for Research on Emotion Sentiment & Social Signals, Istanbul*, pp. 36-39, 2012.

[20] E. Kurtić, B. Wells, G.J. Brown, T. Kempton and A. Aker, "A corpus of spontaneous multi-party conversation in Bosnian Serbo-Croatian and British English", in *Proceedings of International Conference on Language Resources and Evaluation* (*LREC), Istanbul*, pp. 1323-1327, 2012M. Ball, "On percussives", *Journal of the International Phonetic Association,* vol. 28, pp. 95–98, 1998.

[21] S. Fuchs and B. Rodgers, "Negative intraoral pressure in German: Evidence from an exploratory study", *Journal of the International Phonetic Association*, vol. 43, no. 3, pp. 321-327, 2013.

[22] M. Ball, "On percussives", *Journal of the International Phonetic Association,* vol. 28, pp. 95–98, 1998.

[23] A. Marchal, "Des clics en français?" *Phonetica,* vol.44, pp. 30–37, 1987.

[24] J.J. Ohala, "A probable case of clicks influencing the sound pattern of some European languages", *Phonetica*, vol. 52, pp.160–170, 1995.

[25] S. Fuchs, L.L. Koenig and R. Winkler, "Weak clicks in German?" in *Proceedings of 16th ICPhS – International Congress of Phonetic Sciences, Saarbrücken*, pp. 449-453, 2007.

[26] A.P. Simpson, "Acoustic and auditory correlates of non-pulmonic sound production in German", *Journal of the International Phonetic Association*, vol. 37, no. 2, pp. 173-182, 2007.

[27] J. Trouvain, "Laughing, breathing, clicking – The prosody of nonverbal vocalisations", in *Proceedings of the Conference on Speech Prosody (SP7), Dublin*, pp. 598-602, 2014.

[28] A. Cutler, "Learning about Speech", Keynote given at *Interspeech 2014, Singapore*, 15 Sep 2014. (Available online at http://www.superlectures.com/ interspeech2014/ keynote-1-isca-medalist Last access on 2016-03-20).

[29] S. Kousidis, T. Pfeiffer, Z. Malisz, P. Wagner and D. Schlangen. "Evaluating a minimally invasive laboratory architecture for recording multimodal conversational data", in *Proceedings of the Interdisciplinary Workshop on Feedback Behaviors in Dialog*, pp. 39-42, 2012.

[30] M. Żygis, and Jaye Padgett. "A perceptual study of Polish fricatives, and its implications for historical sound change", *Journal of Phonetics*, vol. 38,.no. 2, pp. 207-226, 2010.

[31] Chodroff, E. and Wilson, C. (2014). Burst spectrum as a cue for the stop voicing contrast in American English. The Journal of the Acoustical Society of America, 136(5), 2762-2772.

[32] Forrest, K., Weismer, G., Milenkovic, P., & Dougall, R. N. (1988). Statistical analysis of word-initial voiceless obstruents: preliminary data. The Journal of the Acoustical Society of America, 84(1), 115-123.