



An overview of discourse clicks in Central Swedish

Margaret Zellers

ISFAS, Kiel University, Germany

mzellers@isfas.uni-kiel.de

Abstract

Clicks, ingressive stop sounds produced on a velaric airstream, have been shown to be used for discourse-pragmatic purposes in a number of languages where they are not phonemic. The current work investigates clicks in spontaneous Central Swedish conversation, with a particular focus on their phonetic context and their functionality in discourse. Clicks used to take up a conversational turn or extend a previous one are frequent in Swedish, as are clicks used during a word search or in backchanneling. Unlike reports for some other languages, clicks as markers of stance arise only rarely. While different click functions tend to have different phonetic features surrounding the clicks, the clicks themselves do not appear to be phonetically modified for different functions. This study contributes to the characterization of speech features which do not belong to the linguistic phonological system as such, but which still convey pragmatic meanings in conversational settings.

Index Terms: clicks, Swedish, conversation, spontaneous speech, discourse organization

1. Introduction

Clicks can be defined phonetically as stop sounds which are produced on an ingressive velaric airstream and which tend to have a relatively loud burst [1, 2]. While they can be used phonemically in some Khoisan and Bantu languages, in many other languages of the world, including many European languages, they can be used with a pragmatic or discourse function [3]. When they have these functions, clicks can thus be categorized among what [4] has described as “liminal signs”, having flexibility in their use due to this borderline status. This liminality allows the argument that discourse clicks may be produced unconsciously, proposed by [5]. While phonemic clicks can have a variety of places of articulation, discourse clicks tend to be produced lingually with a dental to alveolar place of articulation, and occasionally with a lateral release [3, 6]. An example of such a discourse click is shown in Figure 1.

In languages in which clicks are not used phonemically, the function of clicks has sometimes been described as communicating a negative assessment or some other negative valence. One particular proponent of this view is [7], who defines the fundamental meaning of clicks in his data as “dissatisfaction”, with compositionality of meaning possible based on the other phonetic features that arise alongside the click. [8] also reports that the majority of clicks found in his German data could be associated with a negatively-valenced assessment. However, while marking of negative stance is clearly a possible function of clicks, a review of the literature suggests that their functions are much broader. [9] reports that in Modern Hebrew (and in the English speech of some American Orthodox Jews), clicks can be used as a general hesitation marker in addition to signalling a negative stance or launching a self-repair. [10] reports on the use of clicks at points of disjunction, e.g. the launching of a new topic or other type of sequence, in both British and

American English; this can be expanded to include the launch of a new turn (i.e. selection of self as the next speaker through clicking) [6, 11], and similar findings have been reported for Russian [12]. [13] also reports on the use of clicks for initiating self-repair in Mandarin Chinese, but these repairs are described as “action repairs”, and may thus be categorized alongside the sequence-launching clicks. Clicks produced when the speaker is conducting a word search (perhaps similar to [9]’s hesitation clicks in Hebrew) have been reported for English [6] and Peninsular Spanish [14].

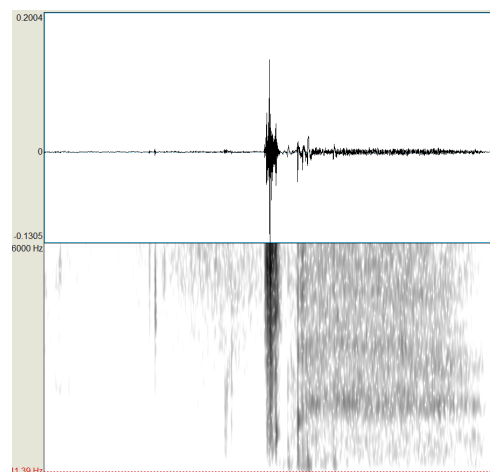


Figure 1: *Waveform and spectrogram of a click, preceded by silence and followed by an audible inbreath, produced by speaker 09-36L from the current dataset.*

Although a set of common discourse functions of clicks have been identified, cross-linguistic differences do appear to arise in the relative frequency or importance of these functions. [6] reports that most of the clicks in his data are turn-initial, with the bulk of the remaining clicks arising in turn-medial position; functions reported in this study are the launching of a new sequence in conversation, and the marking of stance (positive or negative), though no frequency counts are given for these functions. [12] reports new sequence launching and stance marking as functions for clicks in Russian as well, but adds to the list the use of clicks for competitive turn uptake; again, no frequencies are given. [8], on the other hand, reports for his German data that the majority of clicks arise in the context of backchannels, with new turn launches accounting for only about 13% of the tokens. [14, 15] report for Peninsular Spanish that the most frequent functions arising for clicks in their data are either word search, or else a word search combined with stance-taking. Additionally, all of these studies report wide individual differences in clicking, with some speakers clicking often and others not at all.

Given the wide variety of pragmatic clicking behavior ob-

served across these languages, it is valuable to pursue this question for other languages. The current study investigates clicking in Central Swedish. As a North Germanic language, Swedish has structural similarities with the West Germanic languages German and English, but it remains to be seen whether there is also similarity in the clicking behavior of speakers of these languages. This study explores the functions of clicks and their phonetic contexts in spontaneous conversations in Swedish, providing a characterization of the main pragmatic functions and the phonetic constellations that help achieve them.

2. Methodology

2.1. Data

The data used for this study were drawn from the Spontal Corpus of spontaneous Swedish [16]. The corpus consists of two-party conversations between native speakers of Central Swedish, with no prescribed topic of conversation in the portion of the data used. The corpus consists of audio, video, and motion capture data, but only the audio data was used for the current study. Since the audio was recorded using head-mounted microphones, channels are separable and it is possible to carry out acoustic analysis even on speech produced in overlap.

Orthographic transcriptions of portions of the corpus exist, which include generic transcriptions of phonetic features including clicks. An automatic search was made of the orthographic transcriptions to obtain an estimate of the number of clicks available in each conversation. On the basis of this search, the six conversations with the highest number of clicks were chosen for the analysis. The conversation IDs, number of clicks per speaker, and rate of clicks per speaking time, are shown in Table 1, for a total of 303 clicks. There are 12 unique speakers (6 male, 6 female).

Table 1: *Number of clicks in transcribed portions of the six conversations used. The rate of clicks per minute of talking time, calculated as the total transcribed non-silent time for the speaker, is given in parentheses.*

Conversation	Clicks Speaker L	Clicks Speaker R
09-36	32 (2.2/min)	33 (1.6/min)
09-30	31 (2.2/min)	31 (3.5/min)
09-09	26 (1.6/min)	31 (2.6/min)
09-32	25 (4.3/min)	21 (2.4/min)
09-23	16 (3.3/min)	29 (5.4/min)
09-22	25 (3.3/min)	3 (1.0/min)

2.2. Annotations

All annotations were carried out by the author using Praat [17]. Clicks were identified on the basis of the orthographic transcriptions as well as by close listening. This meant that in some cases, clicks that occurred in the orthographic transcriptions were not included as clicks in the current study, or clicks that were not identified by the original transcriber were included. During the annotation phase, no attempt was made to separate between clicks which were produced intentionally versus those which were articulatory epiphenomena. Except for one token which was produced bilabially (and in isolation), the auditory effect of all the tokens indicated a dental-to-alveolar place of articulation, as reported for pragmatic clicks in other languages (see Section 1).

Table 2: *Labels for phonetic features preceding and following clicks.*

Label	Description
s	Silence with duration of at least 120 ms (cf. [19])
i	Audible inbreath
o	Audible outbreath
t	Speech
f	Non-lexical filler particle (e.g. “ehm”)
w	Non-speech sound (e.g. audible swallow)

Table 3: *Function categories.*

Label	Description
Turn uptake (up)	At start of turn chunk after other speaker
Turn continuation (co)	At start of turn chunk, same speaker
Search (se)	Word (or concept) search
Backchannel (bc)	Non-turn-competitive listener response
Stance (st-p, st-n)	Displaying an evaluation (positive or negative) towards a previous turn
Topic/activity change (to)	Turn makes a disjunctive change in direction from previous talk
Repair (re)	Launching a correction to previous (own) speech

In the first annotation phase, clicks were identified in the signal and labelled as to whether they arose at the beginning, middle, or end of a turn chunk (as bounded by Potential Turn Boundaries, cf. [18]), or in isolation. Tokens were classified as arising at the beginning of a turn unit if they occurred at the beginning of a turn chunk following a location where the preceding speech from the same speaker was syntactically or semantically complete, or following speech from the other speaker. Similarly, turn end tokens were identified on the basis that they immediately followed a syntactically or semantically complete unit of speech from the current speaker, or were followed by speech from the other speaker. Turn-medial clicks were those that arose within one speaker’s speech at a location where the speech could not be found to be syntactically or semantically complete. Clicks in isolation were those which were not accompanied by any other speech from the speaker who produced the click (non-lexical filler particles were not classified as speech for purposes of this analysis). Of the 303 clicks identified in the data, 195 were found at the beginning of a turn unit, while 70 were found in the middle of a turn unit; only 3 tokens were found in final position. Of the remaining tokens, 27 were produced in isolation, and 11 could not be classified due to ambiguity of the speech in question.

Once the clicks were identified, the phonetic features for 500 ms on either side of the click were inspected. These were classified as shown in Table 2, roughly following the scheme used by [8]. Multiple phonetic features could be annotated in each context, and no upper limit was prescribed for how many features could arise either preceding or following the click; in practice, no more than 3 features arose together.

In the final annotation pass, either one or two functions were annotated for each click. The annotation scheme was inspired by that used in [14, 15], but the actual functions that arose in the current data were somewhat different, and the annotations were adjusted accordingly. The function annotation categories are shown in Table 3.

2.3. Data extraction

Praat [17] scripts written by the author were used to extract and compile the annotated data for each click, as well as determining on the basis of the orthographic transcriptions whether the click was produced in speech in overlap or in the clear. The mean and maximum intensity of all of the clicks was extracted using Praat's base function, with the minimum F0 set at 60 Hz, and normalized to the speaker's mean intensity.

3. Results

3.1. Phonetic context

For clicks in turn-initial position, the vast majority (152 tokens, 77.9%) were preceded only by silence in the 500 ms window before they were produced. Of the remaining tokens, 9 (4.6%) were preceded by a silence plus an audible inbreath, and 17 (8.7%) were preceded by an audible inbreath filling most or all of the 500 ms window. This contrasts dramatically with the clicks found in turn-medial position, of which only 23 (32.9%) were preceded by complete silence, and 5 (7.1%) by an inbreath. In turn-medial position, clicks tend to be preceded by speech, either filling the whole 500 ms or accompanied by a silence (20 tokens, 28.6%) or fillers, again either filling the whole 500 ms or accompanied by a silence (12 tokens, 17.1%). Audible outbreaths (4 tokens, 5.7%) also occasionally preceded the medial tokens. A maximum of two phonetic context features was identified in the preceding context of any click.

For clicks both in turn-initial position and turn-medial position, the click was most likely to be followed by speech (initial: 82/42.1%, medial: 18/25.7%); a silence followed by speech (initial: 37/19.0%, medial: 16/22.9%); an audible inbreath followed by speech (initial: 17/8.7%; medial: 11/15.7%); or an audible inbreath alone (initial: 29/14.9%, medial: 8/11.4%). Filler tokens also arose, occurring with 25 (12.8%) turn-initial tokens and 7 (10.0%) of turn-medial tokens. In two turn-initial and one turn-medial case, three phonetic features were annotated following the click: silence plus filler plus speech (one turn-initial and one turn-medial) and audible inbreath plus silence plus talk (one turn-initial).

None of the isolated tokens were preceded by anything other than silence or audible breathing (3 inbreaths; 1 outbreath); a speech context was excluded by definition for these tokens. Three of the isolated tokens were followed by a verbal production that was classified as a filler. Seventeen of the 27 tokens (63.0%) were both preceded and followed by silence, making this the most common feature constellation.

3.1.1. Clicks in overlap and in the clear

93 (30.7%) clicks in the dataset were produced in overlap with speech from another speaker, while the rest were produced in the clear. Backchannel clicks were the only function to occur overwhelmingly in overlap; other functions were produced both in overlap and in the clear, with approximately a 1:2 ratio as in the overall data.

3.2. Functions

For 262 (86.5%) of the click tokens, a single function was identified; 33 (10.9%) tokens were annotated with two functions, and 10 (3.3%) tokens could not be assigned a clear function. The four most frequent functions were Turn continuation (111 tokens, 36.6%), Turn uptake (93 tokens, 30.7%), Search (52 tokens, 17.2%), and Backchannel (34 tokens, 11.2%). (Search appeared in combination with Turn continuation in 6 tokens and Turn uptake in 2 tokens, so these tokens are counted double in the preceding.) Topic/activity change (21 tokens, 6.9%), Stance, either positive or negative (17 tokens, 5.6%), and Repair (2 tokens, 0.6%) were rare in the current data.

The functions are relatively constrained by their location; Turn uptake is by definition turn-initial, and Topic/action change also only occurred in this position. Search clicks could only be clearly identified turn-medially, since it became obvious through the context that a search was intended. Turn continuation clicks could occur both in turn-initial and turn-medial position; in the latter, they were distinct from Search clicks in that a particular conversational project was apparently abandoned and a new one launched, but without bringing the first project to a point of completion. Turn continuation clicks without another function in turn-medial position are, however, rare: only 11 tokens were found (3.6%). While Stance tokens could in principle have arisen in any position, in practice, they are about twice as frequent turn-medially as turn initially (12 turn-medial tokens, 5 turn-initial).

All speakers in the sample used clicks for the Turn uptake function, 11 produced Turn continuation clicks, and 10 produced Search and Backchannel clicks (not the same 10 speakers for each). This high incidence must be interpreted with caution, since the speakers were specifically selected on the basis that they were the most frequent click producers. In the current data, however, there is no clear pattern of individual preferences for clicking; that is, it does not appear to be the case that speakers have a specific preferred function for clicking, but rather that speakers either tend to click or not, and if they do click, use clicks for a variety of functions.

All click functions were often preceded by silence, but silence following a click was most likely in Backchannel clicks (i.e. where the speaker was not taking up the turn), arising in 16 of the 34 tokens (47.1%). Search clicks could also be surrounded by silence, but were the type of click most likely to have a filler adjacent; the filler could be either preceding (10 of 52 tokens, 19.2%) or following (4 of 52 tokens, 7.7%), but the click never had a filler on both sides. The most frequent phonetic constellation for Turn continuation clicks was for the click to be preceded by silence and followed either immediately by speech (38 of 111 tokens, 34.2%), or else by an audible inbreath (30 of 111 tokens, 27.0%). A Fisher's Exact test was attempted, but due to the extreme differences in distribution, could not calculate a result.

Although many previous reports about clicks marking stance have indicated that this stance is made (more) clear by the surrounding phonetic material, no obvious patterns arose in the current data, either for stance tokens in general or when they were divided between those showing a positive or a negative valence; this may simply be a result of the few tokens available.

3.3. Intensity

Neither the normalized mean intensity nor the normalized maximum intensity of the clicks showed a statistically significant relationship to the position of the click in the turn, the preced-

ing or following phonetic context, the function of the click, or whether the click was produced in overlap with other talk. Turn uptake clicks had the most variability in their maximum intensity compared to the other three most frequent categories, suggesting that some of these clicks were articulatory epiphenomena rather than intentional productions (cf. comparison of clicks and percussives in [6], but it was not possible to draw a clear boundary on the basis of the intensity alone, and a linear mixed model did not achieve statistical significance.

4. Discussion

The current study investigated clicks produced in spontaneous Swedish conversation. Clicks were found to frequently arise in the contexts of launching a new turn or extending a previously-started one, backchanneling, and word search. The communication of these functions was also supported by a different distribution of phonetic features surrounding the clicks, with fillers arising most frequently adjacent to Search clicks, and inbreaths most likely in the context of Turn continuation. Often, though, clicks arose with stretches of silence of 500 ms (or more) on one or both sides. The acoustic features of the clicks themselves did not appear to vary based on their function; the place of articulation of almost all of the clicks was auditorily identified as being dental-alveolar, and the intensity of the click, which might have provided information about the degree to which it was produced intentionally versus as an epiphenomenon of articulation, did not vary systematically alongside any of the parameters investigated here. Overall, the results emphasize the complex nature of pragmatic signalling using clicks.

4.1. Click functions

Almost all of the functions reported in the literature discussed in Section 1 were unambiguously found in the current data: new sequence/topic launch, word search, turn uptake, backchanneling, stance marking, repair. Thus the set of pragmatic functions carried out by clicking in Swedish does not appear to differ substantially from the pragmatic functions carried out by clicking in other languages. One possible difference is the report for Russian of clicking in turn-competitive contexts [12]. While clicks were found to occur in locations where there was overlapping speech, they were more frequent in cases of non-overlapping speech, with the exception of Backchannels, which are by definition non-turn-competitive. However, one use of a Turn uptake click may be to establish speakership in a context which could become competitive even if it is not yet; that is, to prevent an interlocutor from starting to speak. Some clicks in the current data were followed by silence not only in the 500 ms window measured, but for even longer stretches before speech began. Since long silences are often meaningful in conversation, for example, to indicate an upcoming dispreferred action [20], the fact that these long silences are not treated by the interlocutor as problematic (e.g. through questioning or other follow-ups) seems to indicate that the click has achieved its work of establishing the next speaker.

As reported for English by [7] and [6], and contrary to the findings for Spanish by [14, 15], in the Swedish data, turn-initial clicks were more frequent than turn-medial clicks. It seems likely that the locational distribution reflects the most frequent functions at hand; in Swedish, as in English, marking the launch of a new sequence or sequence part was a highly frequent function, while in the Spanish data reported by [14, 15], word and concept search were much more frequent. However,

before jumping too quickly to cross-linguistic or cross-cultural conclusions, it should be pointed out that the current data, like [6]’s, come from conversation, while [14, 15] used data from interviews. Thus the genre may have played a role in determining which click functions were most relevant for the speakers (cf. also [21]).

4.2. Phonetic context

While there were tendencies for certain phonetic contexts to arise with certain functions (cf. Section 3.2), there were no hard-and-fast rules associating specific phonetic context features with specific functional domains. This finding is not surprising; all of the studies reviewed in Section 1 emphasize the fact that clicks occur in a rich phonetic environment which helps establish their meaning, whether this is proposed to occur in a strictly compositional way, cf. [7], or in a way that is more flexible and open to interpretation. The ability of almost all of the phonetic features considered to arise in almost all of the functions considered means that this work provides evidence more in line with an interpretation-based view. Speakers have a variety of phonetic means at their disposal, including other “liminal signs” [4], and can use them flexibly and perhaps even creatively in the context of a click in order to achieve different communicative goals. Since clicks themselves are phonetically relatively inflexible (cf. discussion in [11]), this context is essential to expand the functionality of these sounds.

4.3. Clicks on the inbreath: a short side note

During the annotation of the phonetic features, it was observed that 3 of the speakers in the sample (2 female, 1 male) produced clicks in the middle of an ongoing inbreath. In total, 8 tokens of this type were observed. Since there were so few, they were not analyzed separately (in the preceding material they are classified as having inbreath both as a preceding and a following feature). A salient feature of Swedish conversation is the ingressive production of short tokens such as *ja* “yes” or *nej* “no”; such ingressives are also reported for Norwegian and Danish [2]¹. If more such click tokens on an inbreath could be identified, it would be interesting to investigate if there are any parallels between these clicks and other ingressive forms.

5. Conclusions

Clicks in Swedish, as in other languages, can be used to achieve a variety of pragmatic functions, including starting new turns or sequences, indicating an ongoing word search, backchanneling, or signaling an evaluation. Their surrounding phonetic environment can be modified to create constellations or configurations of features which help guide an interlocutor to the desired interpretation. The distributions of different click functions may be language-specific and/or genre-specific; thus more research is necessary to provide a clearer characterization of the many pragmatic uses of non-phonemic clicks.

6. Acknowledgements

This work was supported by the project *Sound Patterns and Linguistic Structures at the Transition Space in Conversation* (ZE 1178/3-1), funded by the German Research Council (DFG).

¹[2] also reports such ingressives as occurring in German and English, but on the basis of the author’s informal observations in conversation in all three languages, this phenomenon is vanishingly rare in English or German, while being strikingly frequent in Swedish.

7. References

- [1] P. Ladefoged and I. Maddieson, *The Sounds of the World's Languages*. Oxford, UK: Blackwell, 1996.
- [2] R. Eklund, "Pulmonic ingressive phonation: Diachronic and synchronic characteristics, distribution and function in animal and human sound production and in human speech," *Journal of the International Phonetic Association*, vol. 38, no. 3, pp. 235–324, 2008.
- [3] D. Gil, "Para-linguistic usages of clicks," in *The World Atlas of Language Structures Online*, M. S. Dryer and M. Haspelmath, Eds. Leipzig: Max Planck Institute for Evolutionary Anthropology, 2013. [Online]. Available: <https://wals.info/chapter/142>
- [4] M. Dingemanse, "Between sound and speech: Liminal signs in interaction," *Research on Language and Social Interaction*, vol. 53, no. 1, pp. 188–196, 2020.
- [5] J. Trouvain and Z. Malisz, "Inter-speech clicks in an Interspeech keynote," in *Proceedings of Interspeech 2016, San Francisco, USA*. International Speech Communication Association, 2016, pp. 1397–1401.
- [6] R. Ogden, "Clicks and percussives in English conversation," *Journal of the International Phonetic Association*, vol. 43, no. 3, p. 299–320, 2013.
- [7] N. Ward, "Non-lexical conversational sounds in American English," *Pragmatics & Cognition*, vol. 14, no. 1, pp. 129–182, 2006.
- [8] J. Trouvain, "On clicks in German," *Trends in Phonetics and Phonology: Studies from German-Speaking Europe*, pp. 21–33, 2015.
- [9] S. B. Benor, *Second style acquisition: The linguistic socialization of newly Orthodox Jews*. Stanford University, 2004.
- [10] M. Wright, "On clicks in English talk-in-interaction," *Journal of the International Phonetic Association*, vol. 41, no. 2, p. 207–229, 2011.
- [11] R. Ogden, "Audibly not saying something with clicks," *Research on Language and Social Interaction*, vol. 53, pp. 66–89, 2020.
- [12] L. Paschen, "On clicks in Russian everyday communication," in *Urban voices: The sociolinguistic, grammar and pragmatics of spoken Russian*, N. Richter and N. Thielemann, Eds. Berlin: Peter Lang, 2019, pp. 237–257.
- [13] X. Li, "Click-initiated self-repair in changing the sequential trajectory of actions-in-progress," *Research on Language and Social Interaction*, vol. 53, no. 1, pp. 90–117, 2020.
- [14] D. Pinto and D. Vigil, "Clicks as discourse markers in Peninsular Spanish," *Spanish in Context*, vol. 15, no. 3, pp. 441–464, 2018.
- [15] —, "Searches and clicks in Peninsular Spanish," *Pragmatics*, vol. 29, no. 1, pp. 83–106, 2019.
- [16] J. Edlund, J. Beskow, K. Elenius, K. Hellmer, S. Strömbergsson, and D. House, "Spontal: a Swedish spontaneous dialogue corpus of audio, video and motion capture," in *Proceedings of LREC 2010, Valetta, Malta*, 2010.
- [17] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," <http://www.praat.org/>, 2021.
- [18] M. Zellers, "Prosodic variation and segmental reduction and their roles in cuing turn transition in Swedish," *Language and Speech*, vol. 60, no. 3, pp. 454–478, 2017.
- [19] M. Heldner, "Detection thresholds for gaps, overlaps, and no-gap-no-overlaps," *Journal of the Acoustical Society of America*, vol. 130, no. 1, pp. 508–513, 2011.
- [20] S. C. Levinson, *Pragmatics*. Cambridge, UK: Cambridge University Press, 1983.
- [21] E. Gold, P. French, and P. Harrison, "Clicking behavior as a possible speaker discriminant in English," *Journal of the International Phonetic Association*, vol. 43, no. 3, pp. 339–349, 2013.