THE ACOUSTIC MODELING OF CLICK TYPES

Mats Exter

Heinrich-Heine-Universität Düsseldorf, Germany
exter@phil.uni-duesseldorf.de

ABSTRACT

Clicks are phonetically very complex sounds in every respect. While their articulatory properties are beginning to be quite well-understood, a detailed acoustic model for these sounds is still largely lacking. This paper aims at filling that gap.

Keywords: click, acoustics, model, tongue shape

1. INTRODUCTION

This study focuses on the *abruptly released* click types [5], [10], [7], which in turn can be classified according to the shape of the anterior part of the tongue [9], [8]:

- The so-called alveolar click type [!], IPA: “(post)alveolar”, will here be referred to as *apical*.
- The so-called palatal click type [ǂ], IPA: “palatoalveolar”, will be called *laminal*.

Cross-linguistically, this is to some degree an oversimplification [6]; however, for the point made in this paper, it will serve as a useful point of departure. The decisive difference between the two types, it is assumed here, is in the configuration of the *front* of the tongue, i.e. the anterior part of the dorsum plus the blade. It is a central point made in this paper that the precise place of articulation (e.g. alveolar vs. postalveolar) has very little influence on the acoustic properties of the clicks investigated here.

The model presented here builds primarily on the work of Stevens [11], who has presented valuable insights into the basic acoustic properties of clicks. However, the model is not very detailed and in some respects empirically not quite adequate. It will be attempted here to expand that existing model.

The presentation here will be in the frequency domain; consequently, the output spectrum will be modeled as the combined effects of a source spectrum, a transfer function and a radiation characteristic (the latter will be treated as constant and not considered further). Finally, the output of the model will be compared with empirical data from a language with phonemic clicks, N|uu. The entire modeling and analysis was undertaken in Praat [1].

2. THE SOURCE SPECTRUM

According to Stevens [11], clicks (at least the abruptly released ones) can be characterized as single transients (i.e., produced with a pure volume velocity source) with no additional turbulence noise present in the burst.

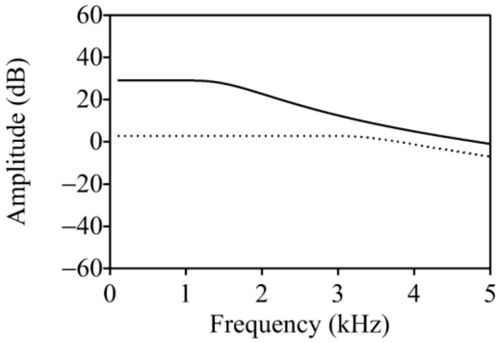
Therefore, as in [11], the estimated source spectrum was calculated piecewise, with the lower frequencies depending on the total volume of air involved in the transient and the higher frequency portions depending on the pressure drop across the (anterior) constriction, and therefore ultimately on the relative magnitude of the change of the volume of the cavity between the two constrictions, $|\Delta V_{cav}|/V_{cav}$. Those two parameters are not independent of each other, however, because the total volume of air in turn depends both on the compliance of the air in the cavity (and hence the initial size of the cavity before expansion, V_{cav}) *and* on the pressure drop across the constriction. For details about the construction of the source spectra, the reader is referred to [11] p. 121-124, which the present work follows exactly.

As Stevens [11] himself points out, he probably underestimated the magnitude of $|\Delta V_{cav}|/V_{cav}$ (for which he uses an estimate of 10%) and thereby also the predicted overall intensity of clicks of the [!] type. Based on the available articulatory and acoustic data (e.g. [5], [12], [7], [9], [8]), the following (very rough) estimates are used here:

- Apical clicks: V_{cav} = 1 cm^3, $|\Delta V_{cav}|/V_{cav}$ = 100%
- Laminal clicks: V_{cav} = 0.5 cm^3, $|\Delta V_{cav}|/V_{cav}$ = 25%

This results in the source spectra given in Fig. 1. As can be seen, the spectra are not dramatically different from each other apart from the overall level. In the output of the model (after application of the transfer function and the radiation characteristic), apical clicks have an intensity that is about ten times that of laminal clicks under the assumptions given above.

Figure 1: Calculated source spectra for an apical click (solid line) and a laminal click (dotted line).



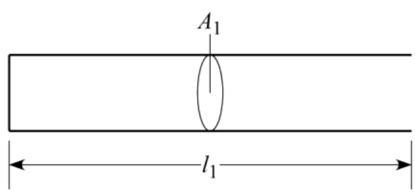
3. THE TRANSFER FUNCTION

Because there is no turbulence noise (due to the very rapid release of the anterior closure), an all-pole transfer function can be assumed for the abruptly released clicks which, according to Stevens [11], can be approximated by the resonances of the vocal tract in front of the posterior closure.

The formant frequencies in this section have been calculated on the basis of a general tube model using the procedures outlined in Fant [2, 3]. Formant bandwidths, on the other hand, have been calculated separately, i.e. not by using Fant's standard equations, but by estimating the influences of the various kinds of acoustic losses individually following Stevens [11]. This is necessary because, as Stevens argues, bandwidths are expected to be quite narrow because there are no losses associated with airflow during the transient.

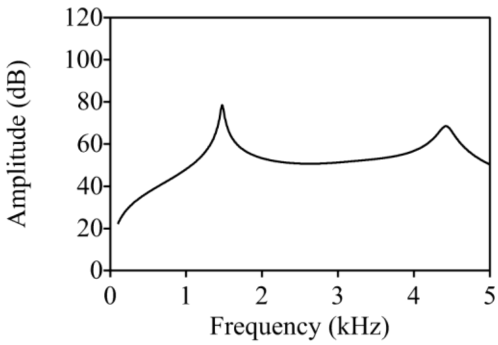
A first approximation to an abruptly released click of the [!] type is given by Stevens [11], who models the vocal tract in front of the posterior constriction as a uniform tube of length l_1 = 6 cm and cross-sectional area A_1 = 1 cm^2 (Fig. 2). Due to the very rapid opening movement, the influence of the anterior constriction is, following Stevens, taken to be negligible in the calculation of the transfer function.

Figure 2: Tube configuration for an (apical, abruptly released) click according to Stevens [11] (l_1 = 6 cm, A_1 = 1 cm^2).



The output spectrum (i.e. the source as discussed above, plus transfer function and radiation characteristic) is shown in Fig. 3; there are two formants, which are the first two natural (quarter-wavelength) resonances of the tube.

Figure 3: Calculated output spectrum for a click based on the tube configuration in Fig. 2. F_1 = 1475 Hz, F_2 = 4425 Hz.



That output spectrum is, however, not yet fully satisfactory because articulatorily, apical clicks are usually produced with a (more or less pronounced) upward-pointing gesture (or retroflexion) of the tongue tip [6]. Such an articulatory gesture leads to the formation of a sublingual cavity which, in turn, introduces an additional pole and a zero (F_R and Z_R in Stevens' terminology) into the transfer function. What is more, the resonance frequencies of the main tube are also altered due to acoustic coupling between the tubes.

Therefore, the original model for [!] was modified in order to allow for a sublingual cavity. Acoustically, the sublingual cavity can be treated as a side branch [11], [4]. The corresponding tube model is given in Fig. 4, and the calculated output spectrum is shown in Fig. 5.

Figure 4: Tube configuration for an apical click with a sublingual cavity (side branch: l_s = 3 cm, A_s = 2 cm^2; length and cross-sectional area of the main tube as in Fig. 2).

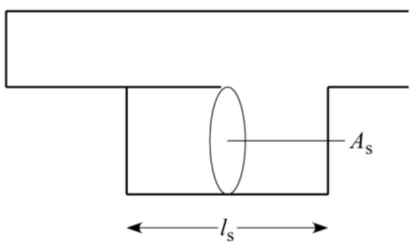
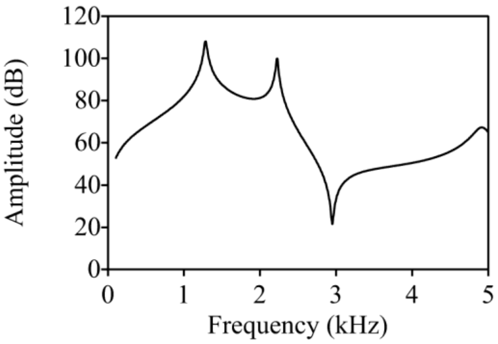


Figure 5: Calculated output spectra for an apical click with a sublingual cavity, based on the tube configuration in Fig. 4. F_1 = 1284 Hz, F_2 = 4917 Hz; F_R = 2225 Hz, Z_R = 2950 Hz.



The consequences of introducing a sublingual cavity are quite profound:

- Lowering of the first natural (quarter-wavelength) resonance of the main tube (= F_1)
- Strengthening of the low-frequency region (below 2.5 kHz) both in amplitude and bandwidth
- Introduction of a zero, thus separating the lower and higher frequency regions more clearly

Having arrived at an improved representation of the apical clicks, we now turn to the laminal clicks. Since (as Stevens [11] has shown) the *size* or *location* of the click cavity as such has no direct consequences for determining the transfer function, and thus the resonance properties of the articulations involved, a different explanation has to be found for the spectral differences between [!] and [ǂ].

It is argued here that the difference lies in the fundamentally different overall tongue configurations between the two types of clicks [9]. Apical clicks, on the one hand, have a *low* tongue center after release, which leads to a configuration that can be approximated by a uniform tube (*after* release), as discussed above. Also, the more-or-less retroflex posture of the front of the tongue introduces a sublingual cavity. In laminal clicks, on the other hand, the front of the tongue forms a very long constriction, which leads not only to a shallower click cavity as seen above, but also to a generally *higher* posture of the center of the tongue. Accordingly, the vocal tract in laminal clicks can be approximated by a *two-tube* configuration as in Fig. 6; the output spectrum of the model for [ǂ] is shown in Fig. 7.

Figure 6: Tube configuration for a laminal click (l_1 = 2 cm, A_1 = 0.3 cm^2; l_2 = 4 cm, A_2 = 1 cm^2).

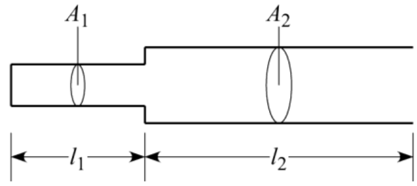
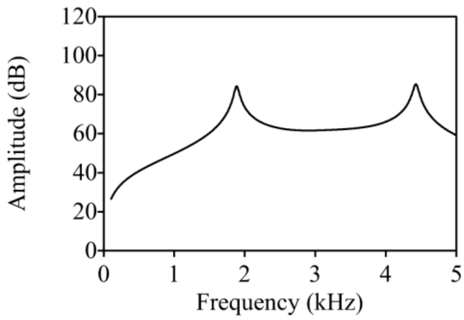


Figure 7: Calculated output spectrum for a laminal click based on the tube configuration in Fig. 6. F_1 = 1885 Hz, F_2 = 4425 Hz.



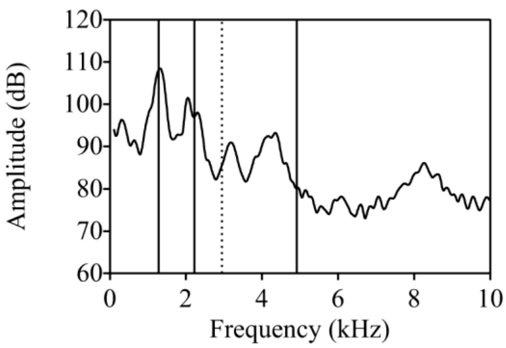
Again, there are two formants, but this time, they are the first natural (quarter-wavelength) resonances of the back and the front tube, respectively. Compared with [!], F_1 is fairly high, and because F_1 and F_2 are quite close together and no zero intervenes between them, they are not as clearly separated from one another, forming one contiguous higher-frequency region in the spectrum.

4. THE EMPIRICAL DATA

How well do the spectra calculated on the basis of this model correspond to the empirical facts? For an assessment of this question, data from 3 female speakers of Nǀuu, a South African Tuu (formerly “Southern Khoisan”) language [9] was compared with the calculations presented above. The speakers were Ouma Katrina Esau, Ouma Anna Kassie and Ouma Hanna Koper; all wished to be acknowledged for contributing to the study.

Ensemble-averaged FFT spectra with a window length of 23 ms were produced for apical alveolar [!] (in *!áà* ‘red hartebeest’) and laminal postalveolar [ǂ] (in *ǂá* ‘someone’; 30 tokens each). The utterances were produced in a phonetically controlled frame sentence (*ná ká*__ ‘I say __’). Fig. 8 shows the averaged spectrum for [!]; superimposed are the calculated frequencies of F_1, F_2, F_R and Z_R.

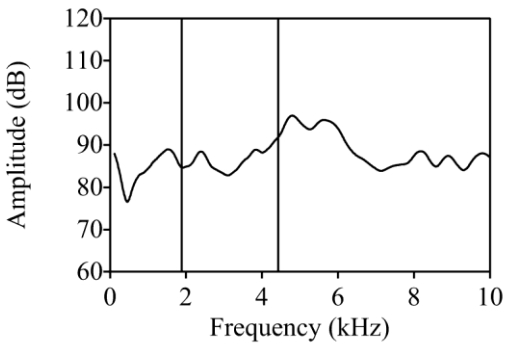
Figure 8: Ensemble-averaged FFT spectrum for [!] (window length = 23 ms, N = 30). Vertical lines show the calculated frequencies of F_1, F_2, F_R (solid lines) and Z_R (dotted line).



Obviously the fit between the model and the data is quite good.

Similarly, Fig. 9 shows the averaged spectrum for [ǂ]. The superimposed lines show the positions of F_1 and F_2 from the model. Here, the fit of the calculated formant values to the data is also convincing, if not quite as good as in the case of the apical clicks. Especially the large difference in level between F_1 and F_2 in the observed data would merit further investigation

Figure 9: Ensemble-averaged FFT spectrum for [ǂ] (window length = 23 ms, N = 30). Vertical lines show the calculated frequencies of F_1 and F_2.



5. CONCLUSION

Two different kinds on influence on the spectra of abruptly released clicks have been discussed in this paper. On the one hand, differences in click cavity size and relative magnitude of click cavity expansion have been identified as influences on the *source spectra* of clicks. Observed differences in intensity ([!]: intense; [ǂ]: less intense) have been plausibly explained in that way; also, the source spectrum enhances the low-frequency regions of the spectrum more in [!] than in [ǂ].

On the other hand, differences in spectral shape (i.e., in the *transfer functions*) have been attributed to differences in the overall articulator configuration after release. It could thus be shown that spectral features such as [grave] and [compact] in apical clicks can be explained by a low tongue position in combination with the presence of a sublingual cavity, whereas features such as [acute] and [diffuse] can be accounted for by a high tongue position (and the absence of a sublingual cavity).

6. ACKNOWLEDGMENTS

I am grateful to the University of Cologne and the German Academic Exchange Service (DAAD) for financial support. I also specifically would like to thank Ouma Katrina Esau, Ouma Anna Kassie, Ouma Hanna Koper, Ouma Griet Seekoei and Grace Humphreys.

7. REFERENCES

[1] Boersma, P., Weenink, D. 2011. Praat: Doing phonetics by computer (version 5.2.25). *http://www.praat.org*

[2] Fant, G. 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton.

[3] Fant, G. 1985. The vocal tract in your pocket calculator. In Fromkin, V.A. (ed.), *Phonetic Linguistics*. Orlando, FL: Academic Press, 55-77.

[4] Jackson, M.T.-T., Espy-Wilson, C., Boyce, S.E. 2001. Verifying a vocal tract model with a closed side-branch. *Journal of the Acoustical Society of America* 109, 2983-2987.

[5] Kagaya, R. 1978. Soundspectrographic analysis of Naron clicks: A preliminary report. *Annual Bulletin, Research Institute of Logopedics and Phoniatrics* 12, 113-125.

[6] Ladefoged, P., Maddieson, I. 1996. *The Sounds of the World's Languages*. Oxford: Blackwell.

[7] Ladefoged, P., Traill, A. 1994. Clicks and their accompaniments. *Journal of Phonetics* 22, 33-64.

[8] Miller, A.L. 2010. Tongue body and tongue root shape differences in N|uu clicks correlate with phonotactic patterns. In Fuchs, S., Toda, M., Żygis, M. (eds.), *Turbulent Sounds: An Interdisciplinary Guide*. Berlin: de Gruyter Mouton, 245-279.

[9] Miller, A.L., Brugman, J., Sands, B., Namaseb, L., Exter, M., Collins, C. 2009. Differences in airstream and posterior place of articulation among N|uu clicks. *Journal of the International Phonetic Association* 39, 129-161.

[10] Sands, B. 1991. Evidence for click features: Acoustic characteristics of Xhosa clicks. *UCLA Working Papers in Phonetics* 80, 6-37.

[11] Stevens, K.N. 1998. *Acoustic Phonetics*. Cambridge, MA: MIT Press.

[12] Traill, A. 1985. *Phonetic and Phonological Studies of !Xóõ Bushman*. Hamburg: Buske.