

COVER PAGE

Document downloaded by @DAEL

Fri Nov 6 16:58:12 2026
For personal use

When automatic English translation is provided, only the original document is authentic.
The EAA cannot be held responsible of any translation error

Bibliographical reference

Acoustic Properties of Tongue Clicks used for Human Echolocation, Rick De Vos and Maarten Hornikx, Acta Acustica **vol. 103** (Number 6), 2017, pp. 1106-1115

DOI
https://doi.org/10.3813/AAA.919138

ACTA ACUSTICA UNITED WITH ACUSTICA
Vol. 103 (2017) 1106 – 1115

Technical and Applied Papers
DOI 10.3813/AAA.919138

Acoustic Properties of Tongue Clicks used for Human Echolocation

Rick de Vos[1)], Maarten Hornikx[2)]
[1)] Royal HaskoningDHV, The Netherlands. rick.de.vos@rhdhv.com
[2)] Eindhoven University of Technology, The Netherlands. m.c.j.hornikx@tue.nl

Summary
In active human echolocation, it is common for visually impaired persons to produce tongue click signals for extracting information from the physical environment around them. The purpose of this paper is to provide a comprehensive dataset of original detailed experimental data on human tongue clicks. The methods used for analysis are described in order to encourage further work to complement the data, which is available online and open for contributions. Series of six different click signals have been recorded from blind persons in an anechoic room. The different click signals all resemble damped oscillations and have in common that the dominant frequency content is located between 1,000 Hz and 8,000 Hz. The standard deviation in click series of the broadband click sound level is between 2.2 and 4.3 dB. The total duration of the clicks, which is mainly determined by the drop-time, ranges between 2.0 ms and 10.0 ms, for a 20 dB decay range. Average standard deviations of 57% in rise-time and 32% in drop-time have been found. The horizontal and vertical directivity of the tongue clicks show a close agreement with literature data from speech signals, but click signals are generally more directive. The minimum and maximum deviations in the 1/3 octave bands from 1,250 Hz to 4,000 Hz are less than 6 dB, for 0 to 120 deg. azimuth, and for −27 to 127 deg. elevation.

PACS no. 43.20.Ye, 43.58.-e, 43.60.-c, 43.60.Cg

1. Introduction

Visually impaired individuals have various complementary means to spatially perceive the physical environment around them, the most important being haptic perception, the white cane, and human echolocation. By using information of the direct produced sound and the acoustic response from the physical environment, i.e. active human echolocation, objects can be perceived. Discrimination can be made between close and distant echolocation. A close object falls within the echo threshold, where the direct sound and echo are not heard separately [1]. Close objects can be detected minutely, i.e. with a spatial resolution of 5 to 10 cm [2, 3], and an angular resolution down to 1.5 degrees [4]. Distant objects as buildings at the opposite side of a street can be detected with a lower spatial resolution. As the source of sound for echolocation, the tongue click was found to be best among several investigated organic excitation signals [5, 6], and its use was also promoted by e.g. Kish and Griffin [7, 8]. For distant echolocation, because of their higher possible amplitude, hand and finger produced pulses and hand-held clickers are considered to be more efficient than oral sounds [6].

Received 2 December 2016,
accepted 17 October 2017.

1.1. Earlier work on tongue clicks in phonetics

Tongue clicks are part of the South African Khoisan language Taa (also known as !Xóō), in which 70 percent of the words start with an oral click sound. Ladefoged and Maddieson [9] provide a detailed description of the sequence of events in the production of tongue clicks based on x-ray tracings of a (native) !Xóō speaker. In short, a body of air is enclosed in the mouth by placing the tip and back of the tongue on the palate. A vacuum is created by a downward movement of the center of the tongue, keeping the tip and back in the same place. Then the tip and sides of the tongue are quickly moved downwards. This results in air rushing into the mouth in order to equalize the air pressure, and a sharp transient sound is produced by this action. Finally, the back of the tongue ends the contact with the soft palate.

It is not obvious how to best categorize oral clicks. In phonetics, oral clicks are categorized based on their linguistic meaning. According to Ladefoged and Maddieson [9], five click types are recognized in language: bilabial, dental, alveolar, palatal and lateral. The articulatory positions and shape of the tongue at the onset, and just before release, determine the (acoustic) characteristics of the click. Kish's Sonic Echolocation review [1] pointed out that many blind individuals use various types of tongue clicks for echolocation. From these, it is acknowledged that the alveolar and palatal clicks are the most common click types. As reported by Ladefoged and Maddieson [9],

© S. Hirzel Verlag · EAA

Document downloaded by @DAEL, 2026-11-06 16:58:14 - personal use only

the waveform of the alveolar click shows damped oscillations with a dominant frequency around 1,200 Hz. The palatal click has less clearly visible oscillations with a dominant spectrum around 3,000 Hz. Fulop *et al.* [10] investigated different tongue clicks types, with different linguistic meaning, on their phonetical and acoustical properties. They found the articulation of the palatal click to be less stable compared to the alveolar click: a palatal click was more often misclassified over other click types (e.g. dental click) and also varies in acoustical and phonetical descriptors. The alveolar click was hardly ever misclassified and also much easier to reproduce. Also, the alveolar click was far more consistent between different speakers.

1.2. Motivation of this paper

In his Sonic Echolocation Review, Kish states [1]: "Theoretical considerations would implicate the click with the sharpest rise-time, shortest duration, greatest intensity, and highest mean frequency as having the greatest utility for echolocation." However, literature on physical properties of tongue clicks (i.e. tongue clicks specifically used for human echolocation) is scarce. In past work, the analysis of physical properties of tongue clicks has not been a primary objective, except in [6]. Often peak frequencies (or ranges), waveforms, duration, and spectral characteristics are reported [3, 6, 11, 12, 13, 14, 15, 16], but the extent of these analyses varies considerably. Rice [3] has reported on some physical properties of tongue clicks (spectrogram and duration), but no information is included about the conditions of the measurements. Rojas *et al.* has presented two examples of tongue clicks used for echolocation, which were recorded in a fitted room [6]. Physical properties. i.e. waveform, spectrum, reproducibility, duration, but also a qualitative judgment (i.e. usability and adaptiveness), is reported [6]. However, the methods for analysis are not fully elaborated, and only limited objective data is provided to support the reported properties. Thaler *et al.* have presented typical waveforms and spectrograms of two individuals, which were recorded with in-ear microphones in an anechoic room [11], but no further acoustic properties are addressed. Schörnich *et al.* have recorded tongue clicks for five subjects, and also report about acoustic properties in [12]. In the acoustic analysis, spectrograms and average values for duration, sound pressure level, peak frequencies and bandwidth are reported. About 3,000 to 8,000 tongue clicks per subject and per experiment were recorded, but no insight is provided on the repeatability of the reported acoustic properties. Also, it is not explained in detail how some of the acoustic properties are determined (i.e. duration and sound pressure level), which makes comparison with literature and repetition of the measurements difficult. Wallmeier and Wiegrebe also have reported physical properties of recorded tongue clicks from sighted and blind individuals: peak frequency range, cut-off frequencies, duration, and sound pressure levels [13, 14]. Multiple clicks were recorded using a headset microphone in an anechoic chamber, but no information concerning repeatability is provided. Thaler and Castillo-Serrano [15] have measured tongue clicks from two experienced blind tongue click based echolocators and 14 sighted subjects with no prior experience with tongue click based echolocation. Two fitted exponential functions to describe the click envelope were used, i.e. a rise-function of the click onset to the click maximum, and a decay-function of the click maximum to the click offset, where the onset and offset amplitude are 95% below the click maximum. The tongue clicks were analyzed on duration, relative RMS intensity, and peak intensity, peak frequency and repeatability. It is not mentioned at what position the clicks were recorded, how many click were analyzed, and no frequency spectra are provided. Thus, although several studies include acoustic properties of tongue clicks used for human echolocation purposes, the analyses are all limited to some extent.

Defining acoustic properties of click signals is important in order to identify the signal properties that favour human echolocation, and can be related to the relevant acoustic cues. For example, the relation between the spectral content of signals and their duration at one hand, and the performance of the human echolocation at the other hand is subject of a companion research paper [16]. The purpose of this paper is to provide a comprehensive dataset of original detailed experimental data on human tongue clicks and its properties. The clicks as produced in an anechoic room are analyzed regarding waveform, duration, spectral content, directivity, and repeatability. The methods for acquiring the desired data are extensively described in Section 2. In Section 3 the click properties as derived from the measurements are discussed. We encourage further work to complement the data in the present paper, which can be found online (see reference in Section 4).

2. Method

2.1. Measurement setup

For every subject, individual multichannel sound measurements have been performed in the semi-anechoic chamber of the Institute of Technical Acoustics (ITA) in Aachen. The measurement setup consisted of a spherical microphone array, with a chair on a computer controlled turntable in its center. The floor of the chamber was covered with a layer of 250 mm PU-foam wedges. The subject was seated such that its head was in the center of the microphone array, which was at 1.8 m above the ground floor. The subject's head was kept secured using a neck support on the chair. The microphone array was arranged according to the geometry of a pentakis dodecahedron with 32 vertices that were equally spaced throughout the array's surface at a distance of 2.08 m from the center. In each vertex a microphone was positioned, which was partially embedded in sound absorbing foam. The microphones, Sennheiser type KE 4-211-2, are omnidirectional and have a flat frequency response in the frequency range of interest. Only the bottom vertex was omitted in order to provide space for the chair. Figure 1 shows a photo of the microphone array. The microphones were connected to four

Document downloaded by @DAEL, 2026-11-06 16:58:14 - personal use only

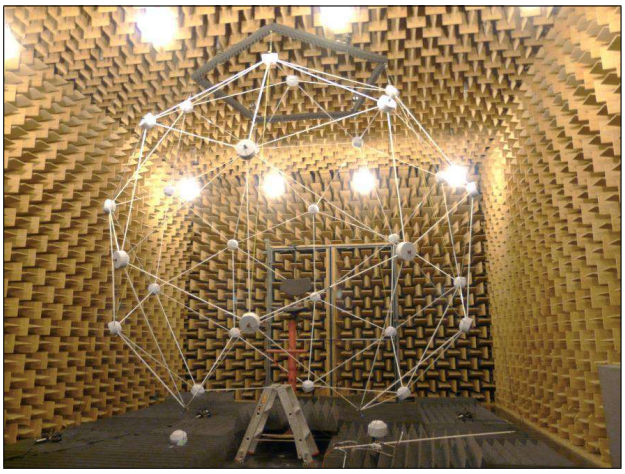


Figure 1. Measurement setup with chair for the subject in the semi-anechoic chamber, ITA Aachen.

synchronized RME-Octamic-D AD-converters, that were connected to an RME Digiface audio device and a PC with multichannel recording software REAPER. Measurements were performed with a sampling frequency of 44,100 Hz, with 24 bits per sample. The measurement setup allowed synchronous audio measurements with 31 microphones.

The geometrical distribution of microphones in the array resulted in 5 microphones per elevational angle, with an angular resolution of 72 deg. azimuth, see Figure 1. The azimuthal resolution of the data was refined by performing separate measurements for 6 different azimuthal angles of the chair. Initially, the chair was oriented such that the subject's direction of view was aligned with 0 deg. azimuth. After the measurement, the chair was rotated with one step of 12 deg. azimuth. This was repeated for 6 measurements in total.

2.2. Subjects' procedure

Four blind subjects participated in the measurements. The subject was introduced to the measurement setup in the semi-anechoic chamber and the procedure was explained. Prior to the measurements, the subject was asked to produce the click types (s)he preferred in most daily situations. The subject was informed when each measurement started and ended. For a measurement the subject was asked to produce one tongue click type at a steady rate of about 1 click (cycle) per second, and keeping its characteristics as constant as possible. Note that in practice tongue clicks are constantly adapted to the needs of the listener, often depending on the acoustic environment. For every click type, measurements of approximately 30 seconds were performed (for 6 different rotations of the chair). Between measurements the subjects were given the opportunity to relax and drink water, ask questions, and practice their tongue clicks.

2.3. Analysis of tongue clicks

Figure 2 shows the spherical coordinates system that is used to express the measurement results. For the horizontal plane the azimuthal angle φ ranges from 0 to ±180 degrees. Because we assume a symmetrical source directivity in the horizontal plane, the directivity data are displayed from 0 to 180 deg. azimuth. In the vertical plane the elevation angle ranges from 0 to ±180 degrees.

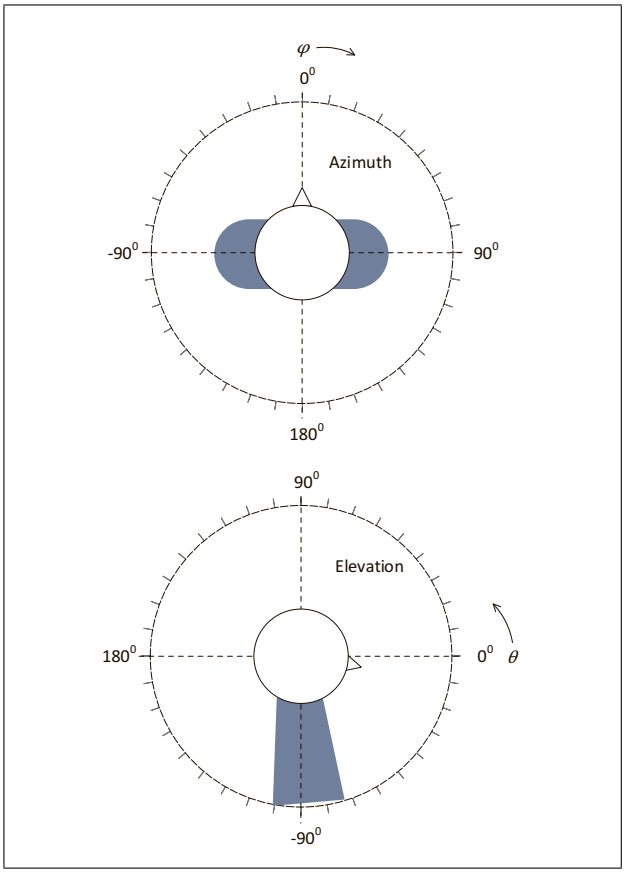


Figure 2. Spherical coordinates system with the origin at the center of the head, (phi) φ in deg. azimuth and (theta) θ in deg. elevation. Both range from 0 to (−)180 degrees, with $\varphi;\theta = 0;0$ deg. as the frontal/viewing direction.

All individual tongue clicks were extracted using a MATLAB script. The upper microphone in the array was used for click detection because of its consistent location (right above the subject, $\theta = 90$ deg. elevation) after each rotation of the chair. The signal of the detection microphone was high-pass filtered (cut-off frequency at 400 Hz). The filtered signal was then condensed, by summation of the energy in blocks of 1000 time samples (23 ms). A tongue click location was assigned in the condensed signal whenever the energy exceeded a +13 dB threshold compared to the minimum energy in the background noise level. For every identified tongue click location, a rectangular time window of 4,000 samples (90 ms) was extracted from the measurement signal from 1,500 samples (34 ms) before to 2,499 (57 ms) after the location marker. For analysis of background noise a rectangular time window of equal size of 4,000 samples was extracted from the measurement signal, 500 samples (11 ms) after the 'click time window', for every detected click. A Fourier transform was applied to every extracted click. The frequency spectrum of each click was determined for 1/3 octave bands.

Document downloaded by @DAEL, 2026-11-06 16:58:14 - personal use only

In this paper, the measured tongue click types are described by:

i. Waveform and normalized spectrogram in frontal direction ($\varphi = 0$ deg. azimuth, $\theta = -11$ deg. elevation);
ii. Normalized 1/3 octave frequency spectrum in frontal direction;
iii. Duration, expressed in rise-time, drop-time and total duration in milliseconds;
iv. Relative directivity in the horizontal plane (azimuth) for 1/3 octave levels and total A-weighted level;
v. Relative directivity in the vertical plane (elevation) for 1/3 octave levels and total A-weighted level.

The standard deviations of the results shed light on the repeatability of the click properties. The click spectra for ii) and were normalized to the total sound pressure level of the reference microphone, which was positioned at -11 deg. elevation and 0 deg. azimuth, because no microphone was positioned at 0 deg. elevation in the spherical microphone array. The energetic averaged spectrum and arithmetic standard deviation were calculated per receiver position as well. This was repeated for every rotational position of the chair. In order to compare data from different rotational positions of the chair for the directivity analysis, all 1/3 octave levels were first normalized to the detection microphone position, because of its consistent location for different rotations.

For the horizontal relative directivities of iv), only microphone positions at -11 deg. elevation are used. Sound levels are related to the level of the reference microphone (0 deg. azimuth, -11 deg. elevation) at the same frequency band and elevational angle. For the vertical relative directivities of v), sound levels are also related to the reference microphone.

For the frontal direction, the normalized 1/3 octave band level $L_{\mathrm{ref},x,j}$ in dB, and the total level and A-weighted level, $L_{\mathrm{ref},x}$ and $L_{\mathrm{A,ref},x}$, for tongue click type x are computed using

$$L_{\mathrm{ref},x,j} = 10\log_{10}\left(\frac{1}{n}\sum_{m} 10^{L_{\mathrm{ref},x,j,m}/10}\right) - 10\log_{10}\left(\frac{1}{n}\sum_{j}\sum_{m} 10^{L_{\mathrm{ref},x,j,m}/10}\right), \tag{1a}$$

$$L_{\mathrm{ref},x} = 10\log_{10}\left(\sum_{j} 10^{L_{\mathrm{ref},x,j}/10}\right), \tag{1b}$$

$$L_{\mathrm{A,ref},x} = 10\log_{10}\left(\sum_{j} 10^{[L_{\mathrm{ref},x,j}+W_A]/10}\right), \tag{1c}$$

where $L_{\mathrm{ref},x,j,m}$ is the sound pressure level at the reference microphone, j are 1/3 octave bands indices for bands ranging from 400 Hz to 20,000 Hz, m is the click number in the measurement, n is the total amount of clicks in the measurement, W_A = A-weighting for 1/3 octave band j. The standard deviation of $L_{\mathrm{ref},x,j}$ is calculated arithmetically. The different tongue click types are given in Table I.

For computing the relative directivities of iv) and v), the average relative (normalized) level, $L_{x,j}$ in dB, and the A-weighted level, $L_{x,\mathrm{A}}$ in dB, for tongue click type x is computed using

$$L_{x,j}(\varphi,\theta) = 10\log_{10}\left(\frac{1}{n}\sum_{m} 10^{L_{x,j,m}(\varphi,\theta)/10}\right), \tag{2a}$$

$$L_{x}(\varphi,\theta) = 10\log_{10}\left(\sum_{j} 10^{[L_{x,j}(\varphi,\theta)-L_{\mathrm{det},x,j}-L_{\mathrm{ref},x,j}]/10}\right), \tag{2b}$$

$$L_{x,\mathrm{A}}(\varphi,\theta) = 10\log_{10}\left(\sum_{j} 10^{[L_{x,j}(\varphi,\theta)+W_A]/10}\right), \tag{2c}$$

where $p_{x,j,m}$ is the sound pressure at θ deg. elevation and φ deg. azimuth, $L_{x,j}$ is the average sound pressure level in dB in 1/3 octave band j in direction (φ,θ), $L_{\mathrm{det},x,j}$ is the average sound pressure level in dB in 1/3 octave band j at the detection microphone ($\theta = 90$ deg. elevation, i.e. above the participant). The average relative level $\overline{L}_{x,j}(\varphi,\theta)$ over all six click types are computed arithmetically as

$$\overline{L}_{x,j}(\varphi,\theta) = \frac{1}{6}\sum_{x=1}^{6} L_{x,j}(\varphi,\theta), \tag{3a}$$

$$\overline{L}_{x,\mathrm{A}}(\varphi,\theta) = \frac{1}{6}\sum_{x=1}^{6} L_{x,\mathrm{A}}(\varphi,\theta). \tag{3b}$$

To quantify the duration of the measured signals, an envelope of the signals is computed utilizing the Hilbert transform [17]. The real part of the obtained signal $H(p_{x,m}(t))$ is equal to $p_{x,m}(t)$ and it moreover contains an imaginary part such that the envelope of the signal can be computed from its absolute value. The envelope is smooth when the signal $p_{x,m}(t)$ has a narrow band frequency content. To promote the dominant narrow-band frequency range of the click signals, first a second order Butterworth band-pass filter is applied to a measured signal $p_{x,m}(t)$. The cut-off frequencies are determined for every click type, see Table III. The envelope of the signal is now computed as

$$p_{\mathrm{env},x,m}(t) = \left| H\left(p_{\mathrm{filt},x,m}(t)\right)\right|, \tag{4}$$

with $p_{\mathrm{filt},x,m}(t)$ a band-pass filtered click signal. The rise-time $t_{r,DR}$ and drop-time $t_{d,DR}$ for a decay range of DR dB are computed as

$$t_{r,DR,x,m} = \left(t_{\max} - t_{DR,a}\right)_{x,m}, \tag{5a}$$

$$t_{d,DR,x,m} = \left(t_{DR,b} - t_{\max}\right)_{x,m}. \tag{5b}$$

With $t_{\max}$ the time at which $p_{\mathrm{env},x,m}(t)$ has its maximum value, $t_{DR,a}$ the first time sample of $p_{\mathrm{env},x,m}(t)$ and $t_{DR,b}$ the last time sample for which the following equation holds

$$DR > 20\log_{10}\left(\frac{p_{\mathrm{env},x,m}(t)}{p_{\mathrm{env},x,m}(t_{\max})}\right). \tag{6}$$

For the decay range DR, values of -20 dB and -10 dB have been explored in this paper. The total duration of the click signal is calculated by adding the rise-time and drop-time. The process of computing the duration of a signal is made visible in Figure 3, where the filtered signal $p_{\mathrm{filt},x,m}(t)$ is plotted from click A4, see Table I, the envelope $p_{\mathrm{env},x,m}(t)$ of this signal is plotted as well as the computed time instances $t_{-20,a}$, $t_{\max}$ and $t_{-20,b}$.

Document downloaded by @DAEL, 2026-11-06 16:58:14 - personal use only

Table I. General information about subjects and their tongue clicks.

Subject	Gender	Age	Usage of active echolocation?	... with tongue click?	Click type	Self-reported effort
R1	M	68	Hardly ever, mostly passive echol.	Hardly ever	A1	Normal
R2	M	30	Very often	Mostly	A2-1 A2-2	Normal Increased
R3	M	52	Very often	Mostly	A3-1 A3-2	Normal Decreased
R4	M	42	Regularly, mostly passive echol.	Hardly ever	A4	Normal

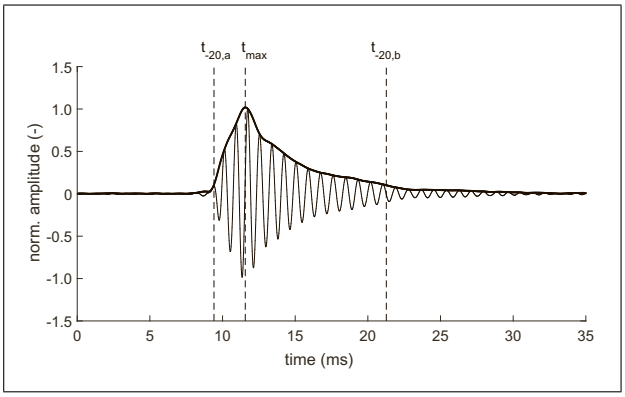


Figure 3. Depiction of the rise-time $t_{r,-20}$ and drop-time $t_{d,-20}$ of a single A4 click (solid thin). Filtered signal p_{filt} (solid thick), envelope p_{env} of p_{filt} computed from the Hilbert transform (dashed), from left to right: $t_{-20,a}$, $t_{\max}$ and $t_{-20,b}$.

3. Results and discussion

Four (visually impaired) subjects participated in the measurements. In our measurements, no echolocation task was given. The measured data therefore are an example of specific click types, and provide information on how well the same click type can be reproduced by an experienced echolocator. Table I gives an overview of the different clicks per subject, as well as some general information about the subjects. Two subjects (R2 and R3) regularly use active echolocation and reported to use two different click types, which both have been documented here. Click types with normal intensity (or effort) would be appropriate in most daily short distance situations. Click A2-2 is reported to be used only when reflecting surfaces would be very far away, or in particular noisy situations. Click A3-2 is reported to be used in quiet situations. Subjects R1 and R4 often prefer to use passive echolocation skills, but also benefit from the active echolocation technique as with self-produced sounds like talking, walking sounds, tapping the white cane and rattling a keychain.

The waveforms and normalized spectrograms of all measured click types are given in Figure 4. Note that these are all single analyzed clicks, as measured with the reference microphone. In Figure 5, the average normalized 1/3 octave band spectra of the different click types are given. The average background noise levels during the measurement sessions are plotted in this figure as well.

The waveforms and spectrograms in Figure 4 show all clicks have a quick rise in amplitude, followed by multiple oscillations that gradually decrease in amplitude. Based on the waveforms alone one can observe different characteristics of the various click types. Especially A4 and A2-2 show clear periodic oscillations, which indicate a specific dominant frequency component. For the other clicks more complex oscillations can be seen, indicating a richer spectrum. Also, different durations of tongue clicks can be observed. From the spectrograms of A1 and A4 it seems that high frequency energy has a shorter duration than the dominant frequency range, but this is not so clear for A2-2, A3-1 and A3-2.

In Figure 5, all click types show dominant peaks between 1,000 Hz up to 8,000 Hz. The acoustic energy of the clicks quickly decreases below 1,000 Hz. The dominant energy is concentrated in a relatively small part of the spectrum. Especially, clicks A2-1, A2-2, and A4 show a narrow-band frequency content, while A1, A3-1 and A3-2 show broader spectra.

In terms of repeatability in the frequency domain, all tongue clicks show standard deviations up to 3 dB in the most dominant 1/3 octave band, see Figure 5. At the remaining frequency bands the standard deviation is somewhat larger, about 5 dB for all click types. This can be explained by the fact that the deviation is computed using arithmetic comparison of all normalized click spectra, with the 1/3 octave band levels per click normalized to its total level. For clicks that have an outspoken narrow-band spectrum, the total click energy is mainly determined by the most prominent 1/3 octave band, which will result in low deviations due to the normalization. Table II shows the deviations with respect to the total relative sound level of the different tongue clicks. The standard deviation of total click levels is between 2.2 and 4.3 dB. The maximal deviation compared to the total normalized average sound level reaches up to 12 dB.

It seems not straightforward to classify the different click types into palatal or alveolar, based on their dominant frequency ranges alone. We think that for the alveolar click, i.e. where the tongue tip is placed further away from the teeth before its release, more clear periodic oscillations occur due to a resonance effect in the mouth cavity (like in A2-2 and A4). For a palatal click this would be less the case as the mouth cavity is much smaller, and maybe less defined, resulting in a broader dominant frequency range

Document downloaded by @DAEL, 2026-11-06 16:58:14 - personal use only

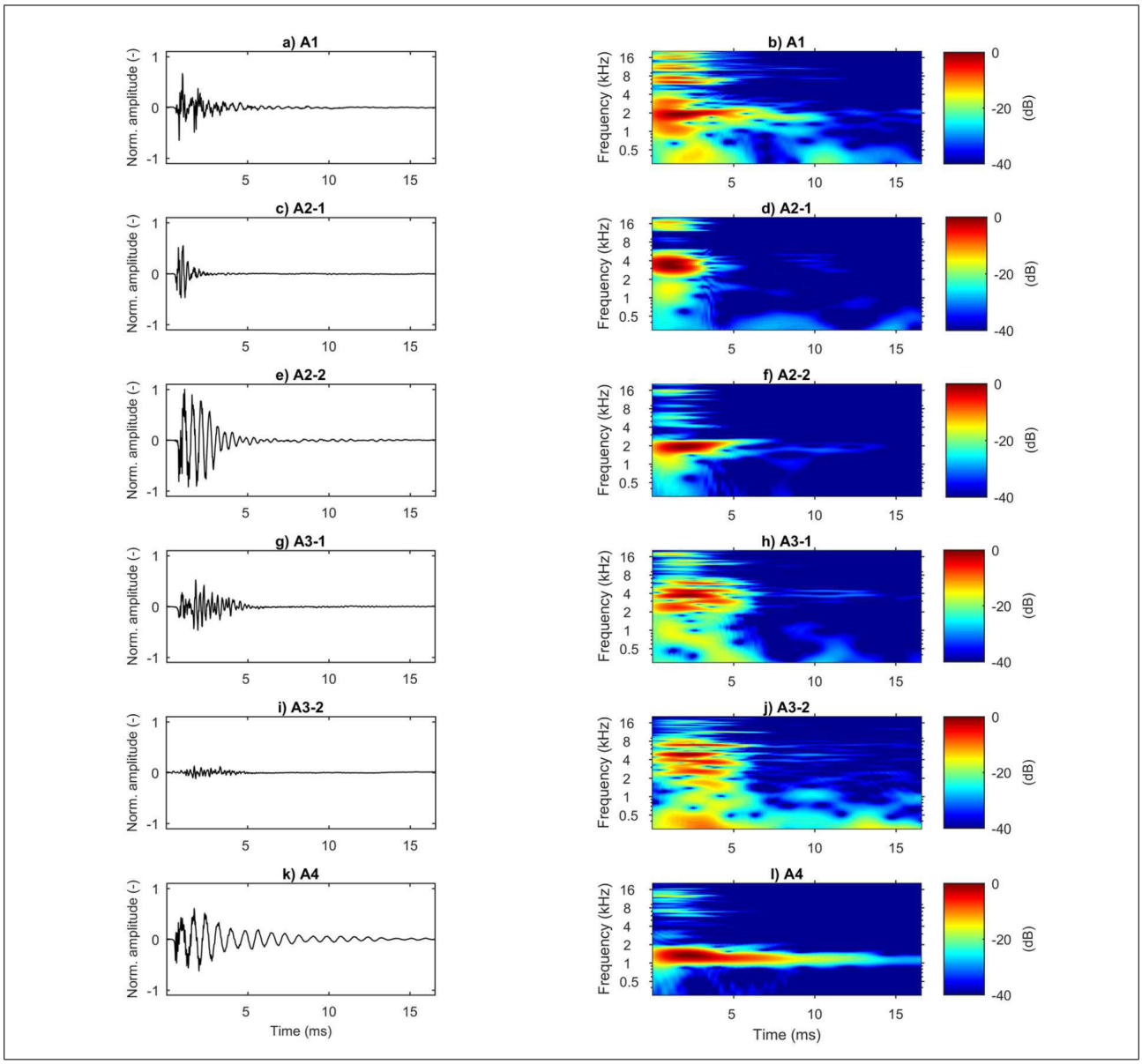


Figure 4. (Colour online) Waveforms (left subfigures) and normalized spectrograms (right subfigures) of measured tongue clicks (one click per click type). All spectrograms are normalized to the maximum value per click type.

(like for A3-1 and A3-2). It seems also very reasonable that mixtures of the two phonetical labeled click types occur within our measured click types.

In Table III the time-domain characteristics of the six different click types are summarized. For two different decay ranges (10 dB and 20 dB) the rise-time, drop-time, and total duration in milliseconds are given, following Equations (4) to (6). The standard deviations, expressed both in ms and as in the percentage of the average, are given in brackets. The relevance of the choice of the decay range in dB depends on its contribution to the envisaged echolocation task and its related echolocation cue. This should be subject of further studies. As the total (A-weighted) sound level of a click signal particularly is determined by the dominant narrow-banded frequency range, and the signal-to-noise ratio reaches 20 dB for all clicks (see again Figure 5), we have been able to determine both decay ranges.

Table II. Deviations of $L_{\mathrm{ref},x}$. Minimum / Maximum: Deviation compared to normalized average.

Click type		Std. dev. (dB)	Minimum (dB)	Maximum (dB)
A1	(n = 43)	4.1	-12.0	6.0
A2-1	(n = 58)	2.8	-6.2	5.2
A2-2	(n = 49)	4.3	-11.9	6.5
A3-1	(n = 22)	2.2	-4.8	5.2
A3-2	(n = 15)	2.6	-4.0	5.0
A4	(n = 28)	3.8	-9.6	5.7

In Table III, the clicks show total durations between 2.0 ms and 10.0 ms, for the 20 dB decay range. These results are in line with other data in literature [1, 9, 12, 13, 14, 15, 16].

Document downloaded by @DAEL, 2026-11-06 16:58:14 - personal use only

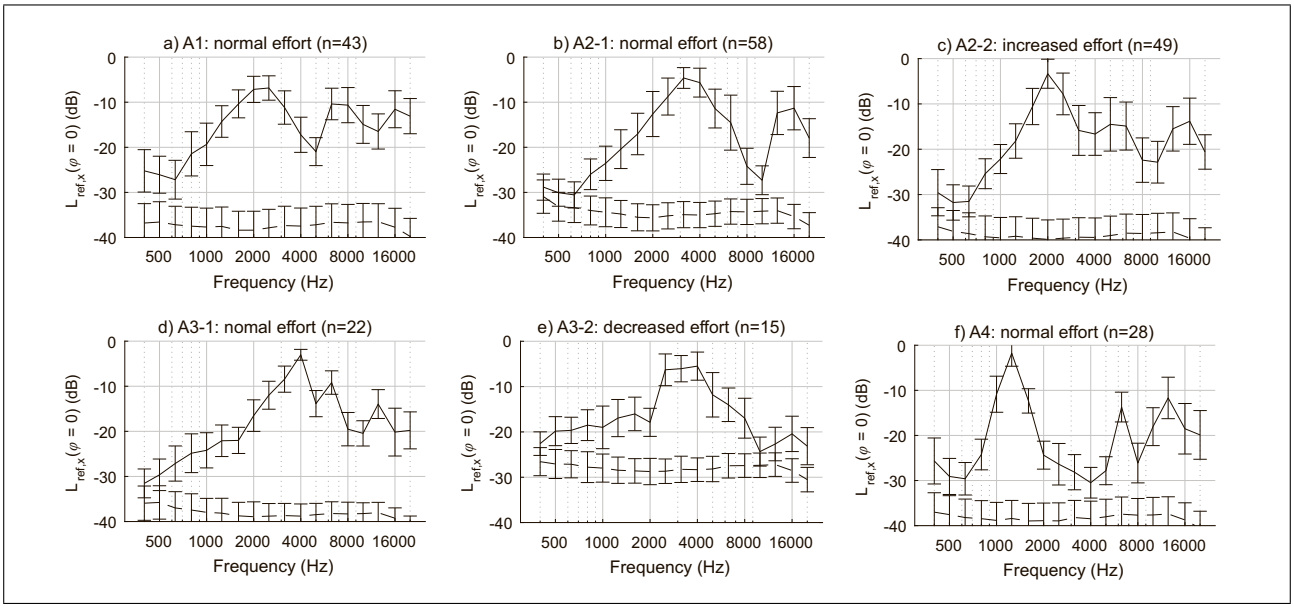


Figure 5. Normalized 1/3 octave frequency spectra of 6 different click types. Solid lines represent average normalized spectra ($L_{\mathrm{ref},x,j}$ according to Equation 1). The dashed lines represent average background noise spectra. The error bars represent the arithmetic standard deviation per frequency band.

Table III. Duration of tongue click types. For every click type the amount of clicks (n) and the dominant 1/3 octave band(s) are given. For the two decay ranges, 10 dB and 20 dB, the table gives the rise-time, drop-time, standard deviation and relative standard deviation in parenthesis, and the total click duration. f_{co}: Band-pass filter cut-off frequencies (Hz); $t_{r/d,x,-10/-20}$: 10 dB/20 dB decay range (ms); τ: Total duration (ms).

Click type	f_{co}	$t_{r,x,-10}$	$t_{d,x,-10}$	τ	$t_{r,x,-20}$	$t_{d,x,-20}$	τ
A1 (n=43)	1250 - 4000	0.9 (1.0 = 117%)	2.1 (1.8 = 87%)	2.9	1.0 (1.2 = 119%)	4.2 (1.9 = 45%)	5.3
A2-1 (n=58)	2000 - 5000	0.4 (0.2 = 65%)	0.7 (0.4 = 57%)	1.1	0.5 (0.3 = 52%)	1.5 (0.5 = 35%)	2.0
A2-2 (n=49)	1250 - 3150	0.7 (0.3 = 41%)	1.4 (0.5 = 33%)	2.1	0.8 (0.3 = 36%)	2.5 (0.7 = 27%)	3.4
A3-1 (n=22)	2500 - 6300	1.8 (1.3 = 70%)	1.6 (0.9 = 58%)	3.5	2.6 (1.2 = 47%)	2.6 (0.8 = 32%)	5.2
A3-2 (n=15)	2000 - 6300	1.1 (0.6 = 54%)	1.7 (1.0 = 57%)	2.8	1.7 (0.7 = 41%)	2.9 (0.9 = 30%)	4.6
A4 (n=28)	800 - 2000	1.3 (0.4 = 34%)	3.9 (0.9 = 23%)	5.3	1.7 (0.9 = 49%)	8.3 (1.8 = 22%)	10.0
Average		63%	52%		57%	32%	

However, in this paper a different way of computing click duration is chosen. A relation between total duration and dominant 1/3 octave band can be expected. Click A4 has clearly the lowest dominant 1/3 octave band of all tongue click types in this paper and also the longest total duration. A larger database of tongue clicks is needed to carry out a proper analysis of the relation between click duration and dominant 1/3 octave band. Most rise-times are shorter than the drop-times. The rise-time indicates how quick the release of the tongue is performed. The drop-time relates to the amount of damping in the resonant system present in the mouth upon releasing the tongue. Clear differences can be found between the results with the 10 dB decay range, compared to the 20 dB decay range. Click types A2-1, A1, and A4 differ in total duration by a factor of about two, which is mainly due to outspoken differences in the drop-times. Click A3-1 clearly shows a longer rise-time compared to the other clicks. This indicates a significantly slower release of the tongue, compared to other click types.

As the regards the repeatability among all click types, the rise-times have a larger relative variability than the drop-times for both the 10 dB and 20 dB decay range. The largest deviation is found for subject R1 (click A1) and the smallest in subject R4 (click A4), who are both less regular users of tongue click based echolocation. Clicks from subject R2 and R3, who frequently use their active echolocation skills with the tongue click, surprisingly show a larger deviation in their preferred click type (i.e. A2-1 for R2 and A3-1 for R3) over their secondary click type, A2-1 and A3-1 respectively. Click A4 has the lowest relative deviations, this agrees with the conclusions in [10]. On average, standard deviations 63% in rise-time and 52% in drop-time can be found for the 10 dB decay range, and these numbers are 57% in rise-time and 32% in drop-time for the 20 dB decay range. Again, we point out the collected data represent tongue clicks that are intended to deviate as less as possible. Therefore, the results we give might not be representative for tongue clicks that are used by echolocators in daily situations.

Document downloaded by @DAEL, 2026-11-06 16:58:14 - personal use only

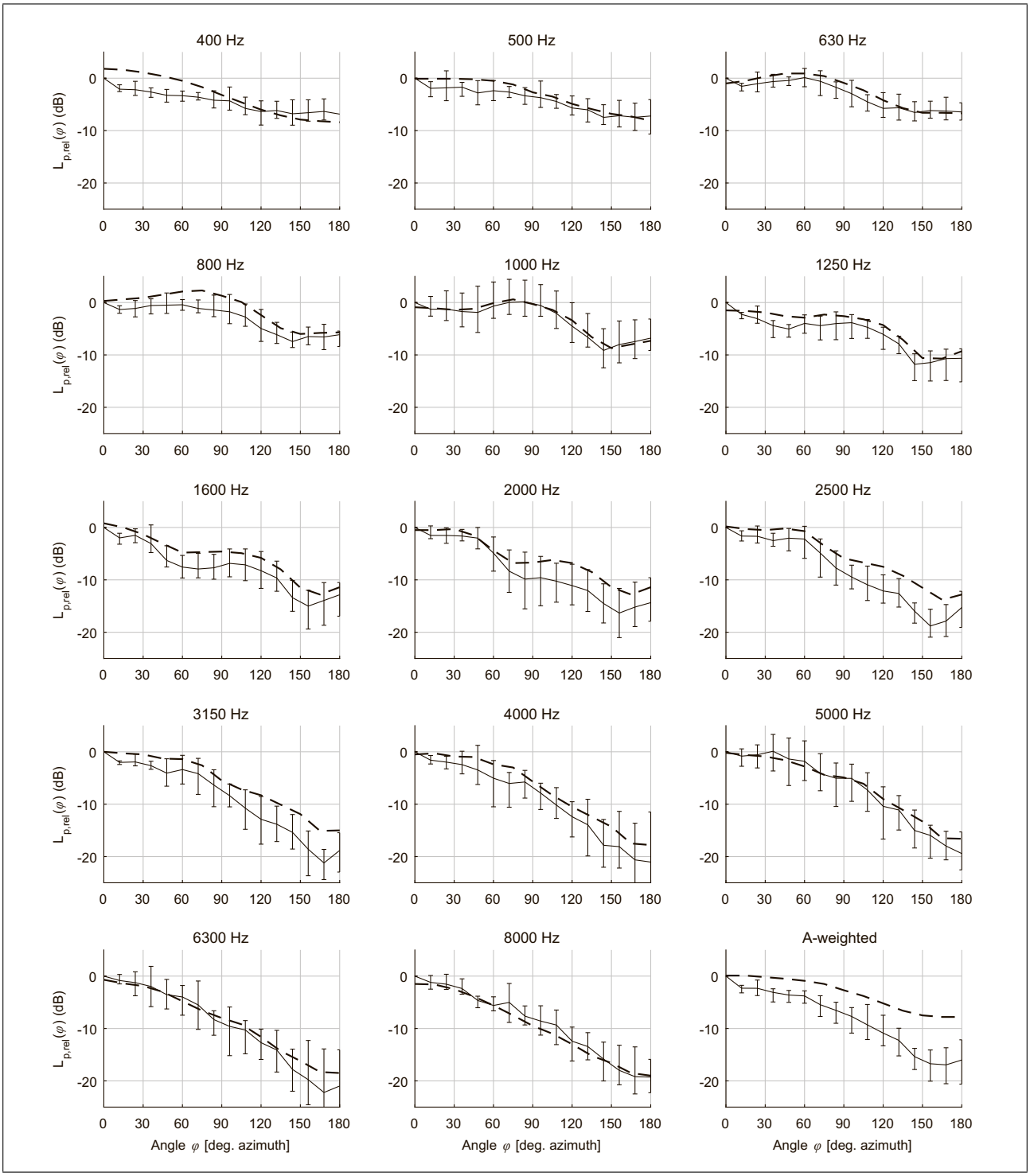


Figure 6. Relative directivity of tongue clicks in the horizontal plane ($\theta = -11$). The solid lines represent average relative directivities $L_{p,\mathrm{rel}}$, in dB, as calculated according to Equations (2) and (3), the arithmetic minimum and maximum deviations (for all click types A1, A2-1, A2-2, A3-1, A3-2 and A4) are expressed with error bars. Dashed lines represent average relative directivity of human speech, by Chu and Warnock [18].

In Figure 6 the relative directivity of tongue clicks as calculated according to Equation (2) is given for the horizontal plane. As the horizontal directivity is considered to be approximately symmetrical, the data from 0 to 180 deg. azimuth are averaged per angle and plotted. Because of the small sample size, i.e. the number of click types, the error bars in the figure represent the minima and maxima of the six measured click types. Furthermore, the directivity data from human speech as reported by Chu and Warnock [18] are plotted in the same figure for comparison. Because our measurement setup did not include a microphone at 0 deg. but at −11 deg. elevation, the speech directivity at eleva-

Document downloaded by @DAEL, 2026-11-06 16:58:14 - personal use only

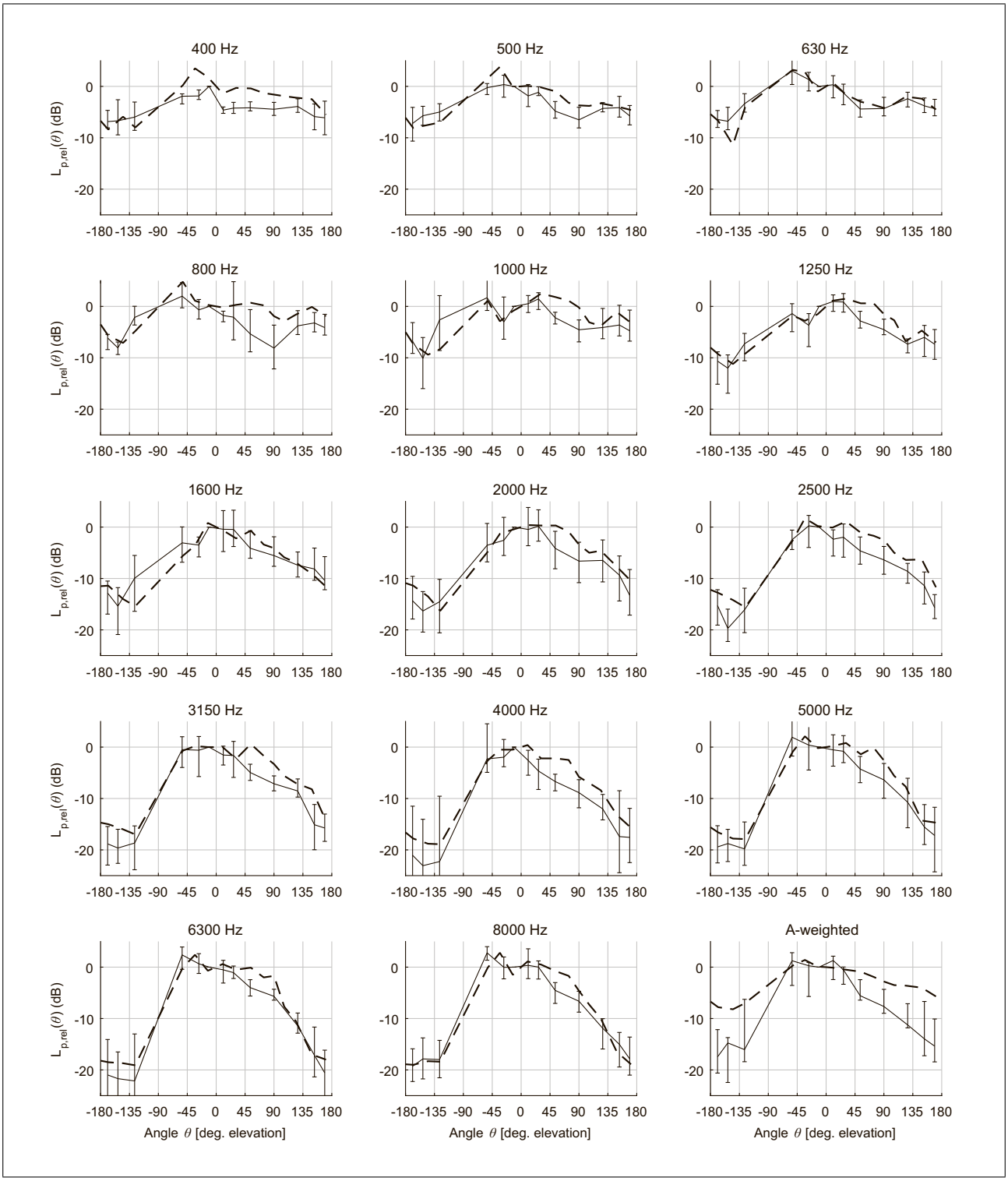


Figure 7. As Figure 6 but for the vertical plane ($\varphi = 0$).

tion −13 deg. from [18] is used for this comparison. In Figure 7 the relative directivity for the vertical plane is plotted. In both Figures 6 and 7 it can be observed that the directivity increases with frequency. Levels are strongest towards the frontal direction. Tongue click directivity is very similar to human speech, but more directive in general. At larger angles in the horizontal plane (Figure 6), the tongue click levels are lower than the speech levels. In the vertical plane (Figure 7) it can be clearly seen that the click levels are more focused toward the lower angles.

Figure 7 makes clear that for high frequencies, the levels at low elevation angles, and also angles behind the participant, are strongly shadowed. There is good agreement between the directivity of tongue clicks and human speech. However, on average tongue clicks are more directive, as the solid lines mostly lie below the speech aver-

Document downloaded by @DAEL, 2026-11-06 16:58:14 - personal use only

age (dashed) in both Figures 6 and 7. Regarding the total A-weighted levels, tongue clicks are more directive than speech, because the frequency content of tongue clicks is dominated by higher frequencies than that of speech (see also Figure 9 from [18]). As can be seen from the error bars the tongue click directivity varies among different click types. The deviations in the 1/3 octave bands from 1,250 Hz to 4,000 Hz are less than 6 dB from 0 to 120 deg. azimuth, and from −27 to 127 deg. elevation. Other factors that may contribute to variation in directivity are the physical position and alignment of the subject in the microphone sphere and body physique of the subject (length, shape, etc.).

4. Conclusions

This paper presents the acoustic properties of tongue click signals produced by blind individuals for human echolocation purposes. The purpose of this paper is to provide a comprehensive dataset of original detailed experimental data on human tongue clicks, which are analyzed regarding waveform, duration, spectral content, directivity, and repeatability. The methods used for analysis are extensively described in order to encourage further work, and allowing to complement the data from the present paper.

Six different tongue click types have been recorded from four blind echolocators. These signals all resemble damped oscillations and have in common that the spectral content is dominated by frequencies between 1,000 and 8,000 Hz. Some signals are dominated by energy in one single 1/3 octave band, whereas other clicks have a somewhat more broadband nature. All click types show a standard deviation up to 3 dB for the dominant 1/3 octave band, and within 5 dB for all studied 1/3 octave bands. The standard deviation of total click sound level is between 2.2 and 4.3 dB. As part of click duration, the envelope of tongue clicks is analyzed on rise-time and drop-time, determined with a 10 dB and 20 dB decay range relative to the maximum click level. For this purpose, a Hilbert transform is applied to the measured signals. The sum of rise-time and drop-time gives the total duration of the clicks, which ranges between 1.1 and 5.3 ms for the 10 dB decay range and between 2.0 and 10.0 ms for the 20 dB decay range. For the 20 dB decay range, average standard deviations of 57% in rise-time and 32% in drop-time have been found.

The directivity of the tongue clicks increases for higher frequencies, with most energy emitted in the frontal direction. The averaged directivity of tongue click signals show close agreement with literature data from speech signals, but tongue clicks are more directive than human speech when the total A-weighted relative sound levels are considered. Among the six studied click types, the minimum and maximum deviations in the 1/3 octave bands from 1,250 Hz to 4,000 Hz are less than 6 dB from 0 to 120 deg. azimuth, and from −27 to 127 deg. elevation.

The dataset presented in this paper is available online at https://dx.doi.org/10.6084/m9.figshare.c.3543852.

References

[1] D. Kish: SONIC ECHOLOCATION: A modern review and synthesis of the literature. Section 5.1: Signal parameters. http://www.worldaccessfortheblind.org/, 2003.

[2] W. N. Kellog: Sonar system of the blind. Science **137** (1962) 399–404.

[3] C. E. Rice: Human echo perception. Science **155** (1967) 656–664.

[4] S. Teng, A. Puri, D. Whitney: Ultrafine spatial acuity of blind expert human echolocators. Experimental Brain Research **216** (2012) 483–488.

[5] J. Rojas, J. Hermosilla, R. Montero, P. Espí: Physical analysis of several organic signals for human echolocation: Oral vacuum pulses. Acta Acustica United with Acustica **95** (2009) 325–330.

[6] J. Rojas, J. Hermosilla, R. Montero, P. Espí: Physical analysis of several organic signals for human echolocation: Hand and finger produced pulses. Acta Acustica United with Acustica **96** (2010) 1069–1077.

[7] D. Kish: Flash sonar program: helping blind people learn to see. www.worldaccessfortheblind.com, 2009.

[8] D. R. Griffin: Listening in the dark: Acoustic orientation of bats and men. Comstock Publishing Associates, Ithaca, New York, 1986.

[9] P. Ladefoged, I. Maddieson: The sounds of the world's languages. Blackwell Publishers Inc., Oxford & Cambridge, MA, 1996. ch. 8.2, 257-259.

[10] S. A. Fulop, P. Ladefoged, F. Liu, R. Vossen: Yeyi clicks: Acoustic description and analysis. Phonetica **60** (2003) 231–260.

[11] L. Thaler, S. R. Arnott, M. A. Goodale: Neural correlates of natural human echolocation in early and late blind echolocation experts. PLoS One **6** (2011) 1–16.

[12] S. Schörnich, A. Nagy, L. Wiegrebe: Discovering your inner bat: Echo-acoustic target ranging in humans. JARO **13** (2012) 673–682.

[13] L. Wallmeier, L. Wiegrebe: Ranging in human sonar: Effects of additional early reflection and exploratory head movements. PLoS One (2014) 1–28.

[14] L. Wallmeier, L. Wiegrebe: Self-motion facilitates echoacoustic orientation in humans. Royal Society Open Science **1** (2014) 1–11.

[15] L. Thaler, J. Castillo-Serrano: People's ability to detect objects using click based echolocation: A direct comparison between mouth-clicks and clicks made by a loudspeaker. PLoS One (2016) 1–14.

[16] R. De Vos, M. Hornikx: Human ability to judge relative size and lateral position of a sound reflecting board using click signals: Influence of source position and click properties. Acta Acustica United with Acustica (2017) submitted.

[17] S. L. Marple: Computing the discrete-time analytic signal via FFT. IEEE Transactions on Signal Processing **47** (1999) 2600–2603.

[18] W. T. Chu, A. C. C. Warnock: Detailed directivity of sound fields around human talkers. RRC NRC Canada, http://irc.nrc-cnrc.gc.ca/ircpubs, 2002.

Document downloaded by @DAEL, 2026-11-06 16:58:14 - personal use only