# ASSESSMENT OF SOUND LATERALITY WITH THE USE OF A MULTI-CHANNEL RECORDER[1]

Anita Lorenc[1], Radosław Święciński[2], Daniel Król[3]

[1] Department of Speech Therapy and Applied Linguistics, Maria Curie-Skłodowska University, Lublin, Poland
[2]Departament of Language and Literature Studies, University of Amsterdam, The Netherlands
[3] Department of Technology, Higher State Vocational School, Tarnów, Poland
[1]trochymiuk@gmail.com, [2]r.j.swiecinski@uva.nl, [3]dankrol@gmail.com

## ABSTRACT

Acoustic analysis of laterality in speech sounds poses numerous obstacles to researchers. Spectral characteristics of such segments vary depending on their phonetic context and shaping of the vocal tract [1]. There are no unambiguous acoustic parameters indicating that a sound is produced laterally. The main alternatives to spectrographic analysis in studying laterality are costly devices, such as electropalatographs (EPG) and magnetic resonance imaging (MRI) scanners.
This article demonstrates how a much more affordable device, the multi-channel recorder, may be used in detecting and assessing laterality in speech. The system records multi-channel audio and calculates spatial coordinates of sound propagation sources, allowing the researcher to establish if the release is central, unilateral or bilateral. This method allows also for evaluating the dominance of the release on the left or right side. It may be used in assessing nasality, too.

**Keywords**: microphone array, lateral articulations, 3D acoustic field distribution, beam-forming.

## 1. INTRODUCTION

Phonetic studies of lateral sounds appear to be more demanding than those of segments that are realized centrally through the oral cavity. Several methods of analysis are used, each having its drawbacks. Acoustic studies, for instance, pose difficulties regarding the choice of methods of measurement, analysis and interpretation of data. After decades of research, no single invariable laterality parameter has been delimited in the acoustic signal and no spectral feature can unequivocally indicate the non-central propagation of the sound wave [2], not to mention distinguishing between uni- and bilateral sounds, let alone the left-right ratio of sound propagation.

This apparent lack of reliability of acoustic analysis makes investigators resort to other instrumental methods, such as EPG [3], [4], MRI [4] and ultrasound imaging [5], [6].
The laterality detection methods enumerated above can be criticized for their cost (MRI, EPG), invasiveness (EPG), unnatural setting of utterance acquisition (MRI, EPG) or inability to record apical articulations (ultrasound). Thus, it appears that the desired device should allow the researcher to acquire data in a non-invasive manner with the recorded person being unrestrained and able to speak naturally. The method described in this article meets all the above criteria.

## 2. BACKGROUND

The system presented here was developed as part of a larger research project devoted to the study of contemporary Polish pronunciation. The project involved an analysis of pronunciation in 20 adult speakers of Polish (10 women and 10 men) who, in the opinion of a team of experts (phoneticians and speech therapists), use the careful style of the standard variety of contemporary Polish. The data discussed in this text were obtained from 4 speakers – two women aged 21 and 24 and two men aged 29 and 46. During the recording sessions, three types of data were acquired from each participant: articulographic, audio and video. Articulographic readouts are not discussed here, however.
The aim of this paper is to present the hardware, signal processing strategies and experimental results of a study whose objective was to develop a technique to detect laterality in the speech signal.

## 3. METHODOLOGY

For the needs of multichannel audio data acquisition a 16-channel microphone-array recorder/processor (MARP-16) was designed and built.

[1] Work described in this paper was supported by grant Nr 2012/05/E/HS2/03770 titled Polish Language Pronunciation. Analysis Using 3-dimensional Articulography with A. Lorenc as the principal investigator. The project is financed by The Polish National Science Centre on the basis of the decision Nr DEC-2012/05/E/HS2/03770.

Unlike similar products on the market, this device is characterized by uncompromising design and construction (The input circuits consisted of low-noise, broadband microphone amplifiers and high-speed successive approximation register (SAR) analog-to-digital converters that are dedicated to measuring equipment. The superiority of the SAR technique over the commonly used sigma-delta technique is presented in the literature [8], [9], [10]. The acquisition and pre-processing of the recorded audio data was performed by a 32-bit floating point digital signal processor (DSP) with the Cortex M4F core. The audio data acquired during the recording sessions were stored on an SDHC/SDXC memory card in the form of 16-channel WAV files. The device was controlled from the main computer equipped in an opto-isolated interface to minimize interference. A circular microphone array was built and fitted with Panasonic WM-61 electret condenser capsules having a linear frequency response.

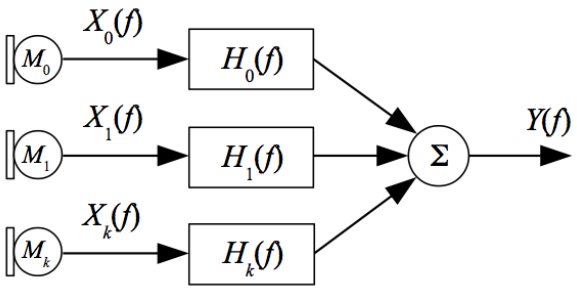
## 4. DATA COLLECTION

During the audio recording sessions, the participants were asked to read out lexical items that were presented consecutively on a screen located at their eye level. Simultaneously, video was recorded by high frame-rate cameras. Referential markers were attached to various locations of their faces so as to provide landmarks used later during analysis.
For the present analysis, recordings of five nouns were selected and examined for each speaker. As our focus was on laterality, each of the recorded tokens chosen for analysis contained the Polish alveolar (sometimes classified as post-dental [7]) lateral consonant /l/. The analysis concerned /l/ that appeared between vowels and constituted the onset of a word-medial stressed syllable (*Malaga*, *kalambur* 'charade', *palarnia* 'smoking room', *inwalida* 'invalid, noun'). Additionally, recordings of the word *tlen* 'oxygen' were analysed as laterality in these tokens appeared not only during the realization of [l] but also during the release of [t].

## 5. SIGNAL SYNCHRONIZATION AND PROCESSING

Following the recording procedure, the audio signals were processed. Combining the 16-channel microphone array with the beamforming method allowed for rendering three-dimensional acoustic fields of the recorded audio data. The procedure was performed in accordance with the filter-sum algorithm [11], [12], [13], [14] in the near field (Fig. 1). The working of the filter-sum algorithm consists in summing the signals from particular microphones and processing them by filters with a finite impulse response (FIR). Beamforming and beamsteering entail determining the appropriate filter weights (see [15] for the application of the delay-sum algorithm in detecting laterality and nasality).

**Figure 1:** The block diagram of the filter-sum beamforming method.




The output signal of the microphone array has the following form in the frequency domain:

$$Y(f)=\sum_{k=0}^{N-1} H_k(f)\,X_k(f)\ , \tag{1}$$

whereas a discrete signal in the time domain is expressed as

$$y[n]=\sum_{k=0}^{N-1}\sum_{m=0}^{M} h_k[m]x_k[n-m]\ . \tag{2}$$

Assuming that the weight of the FIR filter is *M*, we define the sample vector of the input signal as

$$\mathbf{x}_k[n]=[x_k[n],x_k[n-1],...,x_k[n-M]]^T\ , \tag{3}$$

and the vector of filter weights as

$$\mathbf{h}_k=[h_k[0],h_k[1],...,h_k[M]]^T\ , \tag{4}$$

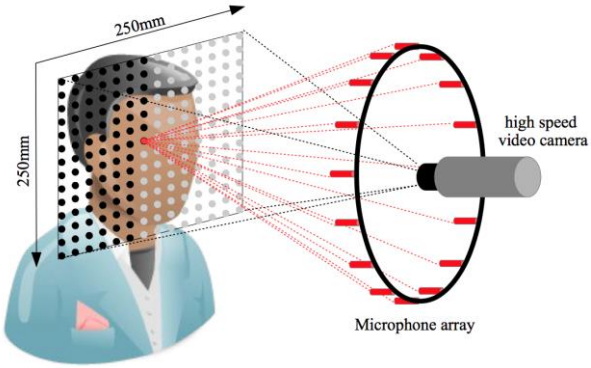
then the filter-sum algorithm output signal may be represented by the equation:

$$y[n]=\sum_{k=0}^{N-1}\mathbf{h}_k^T\,\mathbf{x}_k[n]\ . \tag{5}$$

The array's type and its size were conditioned by the size and shape of the Caarsten's AG500 articulograph cube in which it was installed. The array was placed in the frontal wall of the cube, in front of the speaker's face (cf. Fig. 2).
In order to increase the angular resolution of the beamsteering [13], the sampling frequency in the recorder connected to the microphone array was set to 96kHz, a higher value than used in standard setups. In the study, the microphone array scanned, with the use of the beamforming technique, a 250x250mm square plain with the resolution of 2.5mm. As a result of the specialised application of this, so-called, acoustic camera, at each time point, a

matrix with the dimension 100×100 of the acoustic field distribution was obtained.

**Figure 2:** Scanning of the acoustic field distribution (100x100 points) by the circular microphone array.

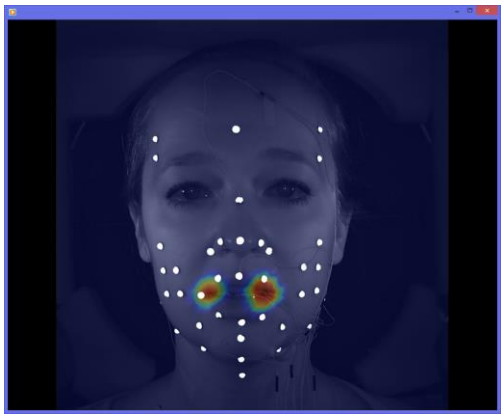



Directionality of microphone arrays such as the one used in this project becomes less well defined in frequencies lower than 1kHz. Thus, to obtain a high spatial scanning resolution, it was necessary to apply narrow-band beamforming. The weights of the FIR filters in the filter-sum block were set to operate in the 1kHz-12kHz frequency range, with simultaneous signal pre-emphasis [16].

The use of beamforming allows for rendering three-dimensional visualizations of the acoustic field distribution, which indicate the active source(s) of the sound pressure when applied to the image of the high-speed camera, as shown in Fig. 3, where acoustic energy concentration is clearly visible in the proximity of the corners of the mouth. This indicates a non-central, lateral, sound release.

Such images were, at a later stage, transformed into graphs showing positioning of the sound source and time, as presented in Fig. 4.

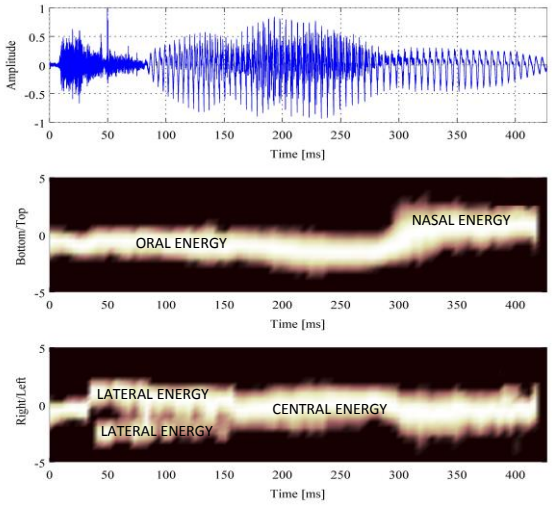
**Figure 3:** Lateral [l] in *tlen* 'oxygen.' The acoustic field distribution image shows bilateral oral sound release.



The graph shows horizontal and vertical distribution of the acoustic field in the vicinity of the nose and lips of the examined participant. The digital signal processing circuit applied Auto Gain, which adjusted each image frame obtained from the acoustic camera. This procedure facilitated observation of subtle differences in the spatial distribution of the acoustic field, independent of the total changes in energy of the input signal. In other words, changes in the amplitude of individual sounds did not mask the changes in spatial distribution.

The vertical distribution of energy in Fig. 4 shows a sudden appearance of acoustic energy at 300ms in the higher region of the graph. It shows that the source of sound propagation has changed and is now higher – 1cm above the zero line. It indicates that the sound ceased to be emitted through the mouth and is released through the nasal cavity.

**Figure 4:** Spatial distribution of acoustic energy in *tlen* 'oxygen.'




The bottom graph in Fig. 4 shows how acoustic energy is distributed when analysing space from side to side (e.g. from the right mouth corner to the left one). Looking at the energy traces, one can notice that energy is projected through the central channel (around the "0" value) until 35ms. Next there appear two parallel energy traces – one above and the other one below the central value. This points to the bilateral nature of the release of the sound.

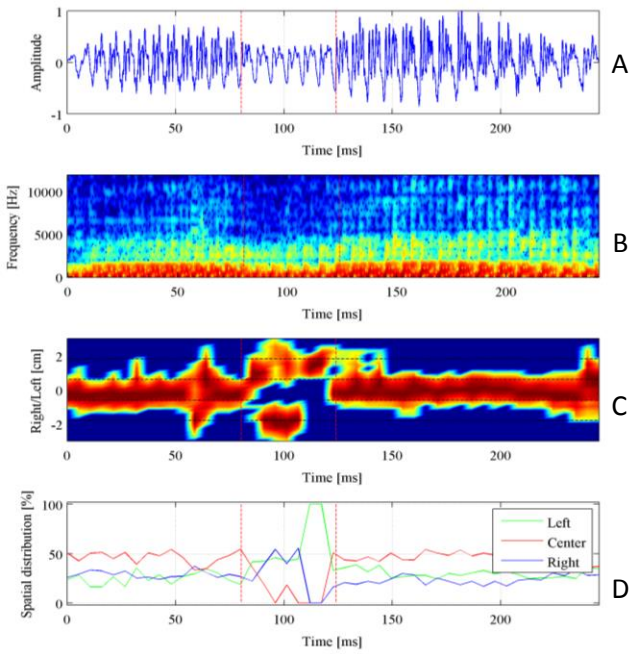
In order to constrain this framework for the assessment of laterality, energy distribution graphs for each token were segmented in the time and space domain and supplemented with additional data, as shown in Fig. 5. Three acoustic energy zones of equal width were distinguished in the lips area (left side, centre, right side). Moreover, boundaries between phonological categories were established in the audio recordings. Such a division of data led to the creation of 9 rectangular cross-sections of the acoustic field distribution as a function of time (Fig. 5c) that enabled spatial analysis of the distribution of

the acoustic energy for individual sounds in a spoken word, as presented in Tables 1–4 for each of the recorded speakers.

## 6. RESULTS

To assess laterality in the recorded words and to test the newly developed system, the obtained data were processed to produce several types of numerical and graphic results. The graph in Figure 5c is representative of the majority of energy distribution plots that were generated. The displayed sound stops to be released centrally after 80ms and dominant energy is present in the right and left channel/lip zone. The graph allows for more observations. Firstly, there appears dominance of the left side during the realization of the lateral. Secondly, the initial part of the realization of /l/ is bilateral and the sound changes to unilateral, which is reflected by discontinuity of the energy path on the right side.

**Figure 5:** Distribution of acoustic energy in [ala].




Acoustic energy was also shown as percentage ratios of its distribution in the three energy zones/channels of the lips area, as is shown in Fig. 5d. The sum of energy in all the three zones at a given point in time constitutes 100%. The energy was calculated as the RMS value (root mean square) in 5ms-long frames.
Looking at the plot, one can notice the prevalence of centrally released sounds (red line). Nevertheless, it dives to very low values, giving in to lateral realizations in the period between 80 and 125ms. It is worth mentioning that laterality as detected by the method described here coincides with the lateral element segmented in the traditional way. It can also be seen in Fig. 5d as the changes in the flow of the centrally released energy occur in the immediate vicinity of segmentation boundaries for the laterals.

The mean values of acoustic energy in individual sounds presented in Tables 1–4 also confirm that the method can be successfully used in identifying the lateral aspect of sounds. The mean results show dominance (marked in bold) of lateral channels in /l/ and of the central channel for the vowels, which is the desired outcome. Moreover, identifying the dominance of either side during the production of /l/, we obtained the following results: ES: right – 3 words, left – 1 word; ZK – Left – 4 words; AK – Left – 2 words, right – 1 word; PT – Left – 3 words, right – 1 word. The dominance was identified whenever a channel was at least 3dB louder than the others.

**Table 1**: Distribution of mean energy (%) in the three areas (left, centre, right side) during the realization of segment /l/ (mean values).

| Speaker | Left | Center | Right |
|---|---|---|---|
| ES female | 27.87 | 23.71 | **48.41** |
| ZK female | **46.35** | 26.92 | 26.73 |
| AK male | **38.22** | 28.45 | 33.33 |
| PT male | **36.60** | 34.81 | 28.59 |

**Table 2**: Distribution of mean energy (%) in the three areas (left, centre, right side) during the articulation of vowels preceding /l/ (mean values).

| Speaker | Left | Center | Right |
|---|---|---|---|
| ES female | 24.26 | **46.20** | 29.54 |
| ZK female | 33.19 | **38.81** | 28.00 |
| AK male | 30.36 | **44.84** | 24.80 |
| PT male | 26.85 | **44.41** | 28.74 |

**Table 3**: Distribution of mean energy (%) in the three areas (left, centre, right side) during the articulation of vowels following /l/ (mean values).

| Speaker | Left | Center | Right |
|---|---|---|---|
| ES female | 29.85 | **47.52** | 22.62 |
| ZK female | 31.53 | **44.85** | 23.63 |
| AK male | 30.31 | **45.31** | 24.38 |
| PT male | 28.66 | **42.16** | 29.19 |

## 7. CONCLUSIONS AND FUTURE WORK

The article presents results of the analysis of the spatial distribution of energy in the acoustic field. The obtained data are very promising for the identification of laterality in speech. The developed method not only allows one to detect laterally released sounds, but it also enables the researcher to make judgements about the dominance of either of the sides in articulation and to quantify this dominance.
Undoubtedly, the results of the method developed here show that it is a non-invasive and effective tool that can be used to objectify the assessment of lateral articulations.

## 8. REFERENCES


[1] Ladefoged, P., Maddieson, I. 1996. *The sounds of the world's languages*. Oxford: Blackwell.

[2] Gubrynowicz, R. 1999. Design and Implementation of Polish Speech Database under the BABEL Project. In: W. Jassem, Cz. Basztura, G. Demenko, K. Jassem (eds.), *Speech and Language Technology*, vol. 3. Poznań: Polish Phonetics Association, 257-275.

[3] Recasens, D. 2012. Coarticulation in Catalan Dark [l] and the Alveolar Trill: General Implications for Sound Change. *Language and Speech* 56(1), 45–68.

[4] Narayanan, S., Alwan, A. 1997. Toward articulatory-acoustic models for liquid approximants based on MRI and EPG data. Part I.The laterals. *Journal of the Acoustical Society of America*, 101, 1064–1077.

[5] Gick, B., Campbell, F., Oh, S., Tamburri-Watt, L. 2006. Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *Journal of Phonetics* 34, 49–72.

[6] Scobbie, J. M., Punnoose, R. and Khattab, G. 2013. Articulating five liquids: a single speaker ultrasound study of Malayalam. In L. Spreafico and A. Vietti (eds.) *Rhotics: New Data and Perspectives*. BU Press, Bozen-Bolzano. 99-124.

[7] Jassem, W. 2003. Illustrations to the IPA. *Journal of the International Phonetic Association* 33(1), 104-107.

[8] Król, D. 2007. Choice of analog-to-digital converters for audio measurements using MLS algorithm. 15th European Signal Processing Conference, EUSIPCO, 3-7 September 2007, Poznań, Poland.

[9] Król, D. 2008. On superiority of Successive Approximation Register over Sigma Delta AD converter in standard audio measurements using Maximum Length Sequences. International Conference on Signals and Electronic Systems, ICSES'08, 14-17 September 2008, Kraków, Poland.

[10] Król, D., Wielgat, R., Potempa, T., Świętojański, P. 2011. Analysis of Ultrasonic Components in Voices of Chosen Bird Species. Forum Acusticum, 26 June - 1 July 2011, Aalborg, Denmark.

[11] Brandstein, M., Ward, D. 2001. *Microphone arrays: signal processing techniques and applications.* Berlin: Springer.

[12] Benesty, J., Chen, J., Huang, Y. 2008. *Microphone Array Signal Processing*. Berlin: Springer.

[13] McCowan, I. 2001. Microphone arrays: a tutorial. http://www.idiap.ch/~mccowan/arrays/tutorial.pdf

[14] Król, D. 2014. Macierze mikrofonowe i głośnikowe. In: Zieliński, T.P., Korohoda, P., Rumian R. (eds.), *Cyfrowe przetwarzanie sygnałów w telekomunikacji: Podstawy, multimedia, transmisja*. Warszawa: PWN, 665-695.

[15] Król, D., Lorenc, A., Święciński, R. 2015. Detecting Laterality and Nasality in Speech with the Use of a Multi-Channel Recorder. *Proceedings of the 40th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Brisbane, Australia.

[16] Loweimi, E., Ahadi, S.M., Drugman, T., Loveymi, S. 2013. On the Importance of Pre-emphasis and Window Shape in Phase-Based Speech Recognition. 6th International conference: Nonlinear speech processing.