DETECTING LATERALITY AND NASALITY IN SPEECH WITH THE USE OF A MULTI-CHANNEL RECORDER[1]

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

[1]Department of Technology, Higher State Vocational School, Tarnów, Poland
[2]Department of Speech Therapy and Applied Linguistics, Maria Curie-Skłodowska University, Lublin, Poland
[3]Department of Phonetics and Phonology, Maria Curie-Skłodowska University, Lublin, Poland

ABSTRACT

Phonetic studies of nasal and lateral sounds pose numerous obstacles to researchers because there are no unequivocal acoustic parameters indicating these types of articulation. Therefore, investigators resort to special systems dedicated to the investigation of nasality and electropalatography (EPG), which is the main accessible alternative to spectrographic analysis in studying lateral sounds. These systems are costly, often invasive and do not examine naturally produced speech.

The present article shows how a multi-channel recorder may be used in detecting non-invasively both nasality and laterality in speech. The described system records video together with multi-channel audio and calculates spatial coordinates of sound propagation sources. Such a combination of audio and video data allows the researcher to establish if the release of an articulated segment is oral, nasal(ized) or lateral.

Index Terms — microphone array, laterality, nasality, 3D acoustic field distribution, beam-forming.

1. INTRODUCTION

Phonetic studies of nasal, nasalized and lateral sounds appear to be more demanding than those of segments that are realized centrally through the oral cavity. Nasality investigators have developed numerous methods of analysis, each having its drawbacks (for a comprehensive review of procedures and techniques used in the assessment of nasality see [1] or [2]). Acoustic studies, for instance, pose difficulties regarding methods of measurement, analysis and interpretation of data. After decades of research, no single invariable nasality parameter has been delimited in the acoustic signal and no spectral feature can unequivocally indicate the lowering of the velum and opening the velopharyngeal port, not to mention the degree of its opening [3]. This apparent lack of reliability of acoustic analysis spawned the development of a range of other instrumental methods. Amongst these, one can enumerate TONAR-based [4] nasometry that analyses parallel signals from two microphones: oral and nasal. The method is used primarily in clinical studies of velopharyngeal disorders and cleft palate speech [5]. Moreover, there is myography [6], [7] with electrodes registering muscle activity, fiberoptic endoscopy with an endoscope inserted into the nasal cavity [8], potentially hazardous X-ray based cinefluorography [9], magnetic resonance imaging (MRI) [10], [9], Velotrace inserted into the nasal cavity [11], electromagnetic articulography (EMA) with a sensor attached to the velum [12], devices with masks equipped in differential pressure transducers (e.g. Rothenberg mask [13]), and other appliances.

Most of the nasality detection methods enumerated here can be criticized for their cost (e.g. MRI, EMA), invasiveness (e.g. endoscopy, myography), unnatural setting of utterance acquisition (e.g. horizontal speaker position during MRI recordings or a mask on the face while speaking), or direct contact of the device with the velum, which may impede its natural movements (e.g. Velotrace, myography, EMA). Thus, it appears that the perfect device should allow the researcher to obtain nasality data in a non-invasive manner with the recorded person being unrestrained and able to speak naturally.

While investigations of nasality offer several instrumental options, electropalatography (EPG) is the only reliable and relatively accessible non-spectrographic way of investigating

[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.

laterality of sound realization known to the authors of the present paper. This technique is rather costly as each new recorded person requires a dedicated electrode-filled artificial palate; the presence of such a palate in the oral cavity results in the production of slightly distorted, unnatural speech.

Both articulatory features discussed here have to be studied separately if one does not want to resort to spectrography. Nasality analysis systems do not allow for the possibility of examining whether the oral release is lateral or not, while electropalatography returns data on linguo-palatal contact and the positioning of the soft palate cannot be scrutinized. Thus, if one needs to investigate instrumentally both nasality and laterality of sounds, they need to resort to two separate devices or systems.

This article, however, presents a system that allows for a simultaneous detection of laterality and nasality – a feature that is not offered by any other device available to phoneticians and speech investigators.

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. Three types of data were acquired from each participant: articulographic, audio and high-speed video. Since articulographic data are not relevant for the estimation of nasality and laterality, they will not be presented here.

The participants were qualified for the study on the basis of specially developed criteria related to linguistic (phonetic, orthophonic, sociolinguistic) and biological (anatomic, functional and perceptual) factors [14]. The qualification interviews allowed for the exclusion of participants with anatomical defects within the articulatory apparatus (e.g. bite malocclusions, dental abnormalities, malformations of the lips, tongue, palate, etc.), disorders of motor control and oral parafunctional habits (such as swallowing or chewing) as well as abnormalities related to hearing.

3. METHODOLOGY

For the needs of multichannel audio data acquisition a 16-channel microphone-array recorder/processor (MARP-16) was designed and built (Fig. 1).

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 [15], [16], [17]. 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.

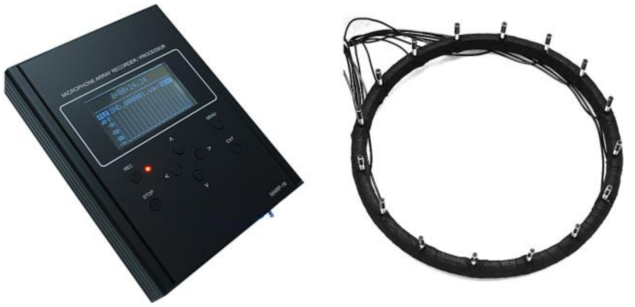

Fig. 1. The 16-channel audio recorder/processor and circular microphone array used in the study.

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 (cf. Fig. 1) was constructed and equipped with Panasonic WM-61 electret condenser capsules having a linear frequency response.

4. DATA COLLECTION

During the recording sessions, the participants were asked to read out 382 lexical items that were presented consecutively on a screen located at their eye level. The list of tokens was constructed on the basis of formal (grammatical category, morphological complexity) and phonological (complete inventory of phonemes, the length of the word and its phonetic structure) criteria. The participant's task was to memorize the word presented on the screen and then say it aloud in the most natural way after the screen turned green. Between some tokens and the green screen there appeared a distractor in the form of a simple mathematical operation (subtraction) whose result was to be said aloud by the participant. The use of such distractors and delayed response aimed at minimizing spelling pronunciation.

5. SIGNAL SYNCHRONIZATION AND PROCESSING

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 delay-sum algorithm [18], [19], [20], [21] in the near field. Fig. 2 presents a block diagram of the delay-sum algorithm which consists in delaying the signal from each microphone by d_k samples and then summing the signal values.

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

$$y(t)=\frac{1}{N}\sum_{k=0}^{N-1}x_k(t-\tau_k)$$

The discreet signal may be represented by the equation:

$$y[n]=\frac{1}{N}\sum_{k=0}^{N-1}x_k[n-d_k]$$

where d_k is a delay expressed as the number of samples at a given sampling frequency f_{pr}.

$$d_k=f_{pr}\tau_k$$

Summing up, the delay-sum algorithm compensates for the time shift in microphone array recorded signals coming from a given distance and direction.

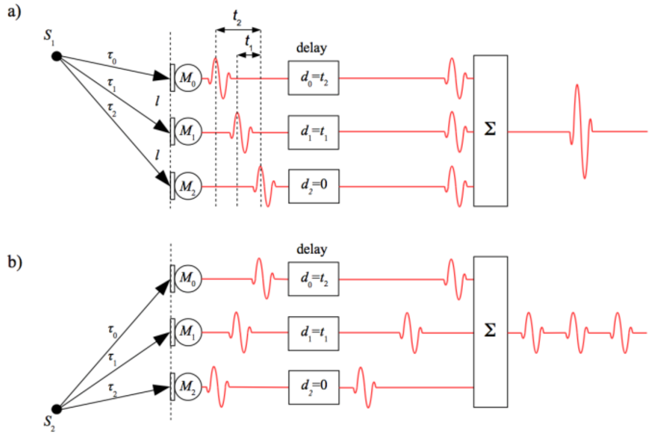

Fig. 2. The block diagram of the delay-sum beamforming method.

Before the beamforming calculations, the signals from each microphone were normalized and pre-emphasized [20], [22]. The array's type and its size were based on 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. 3).

In order to increase the angular resolution of the beamsteering [21], 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 500x500mm square plain with the resolution of 5mm. As a result, at each time point, a matrix with the dimension 100×100 of the acoustic field distribution was obtained.

6. RESULTS

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. 4 and Fig. 5.

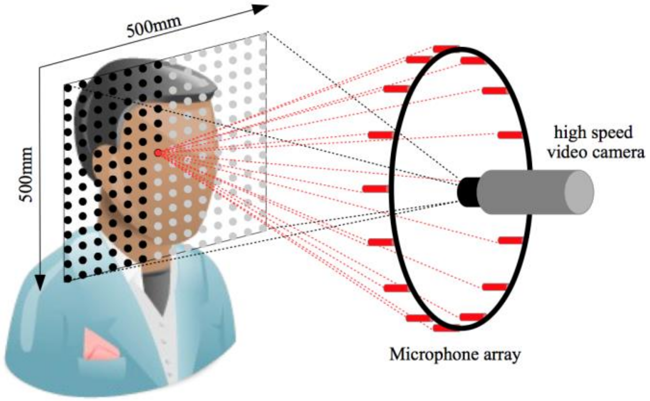

Fig. 3. Scanning of the acoustic field distribution (100x100 points) by the circular microphone array.

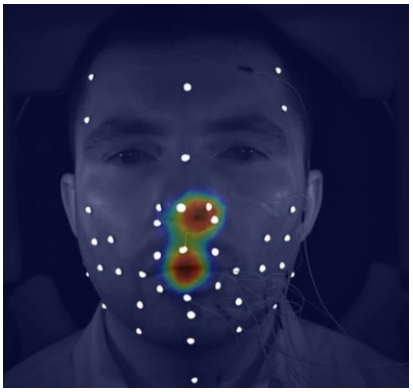

Fig. 4. Nasalized [ɔ̃] in *tobą* (2nd person pronoun, sing. instr.). The applied techniques show that the sound propagates simultaneously from the oral and nasal cavity.

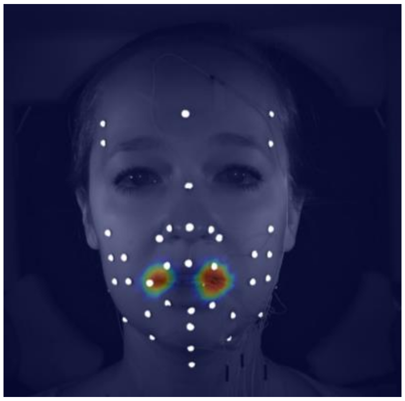

Fig. 5. Lateral [l] in *tlen* (Eng. *oxygen*). The acoustic field distribution image shows bilateral oral sound release.

In addition, it is possible to generate vertical and horizontal cross-sections of the acoustic field distribution as a function of time (Fig. 6 and Fig 7) that facilitate spatial analysis of the distribution of the acoustic energy for individual sounds in a spoken word. The vertical section lets one investigate the existence and level of speech nasality (visible in Fig. 6 as elevation of energy source starting at 300ms and in Fig. 7 between 120 and 290ms). In a similar manner consonant laterality may be scrutinized in the horizontal section (visible in Fig. 6 in the 40-150ms segment as two simultaneous non-central energy streams).

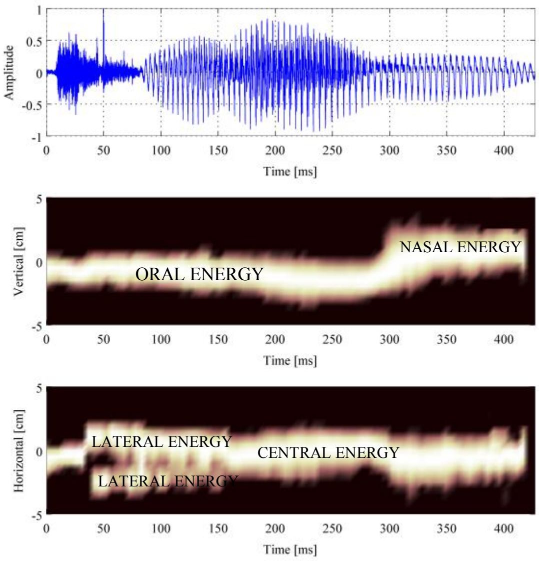

Fig. 6. Spatial distribution of acoustic energy in *tlen* (Eng. oxygen).

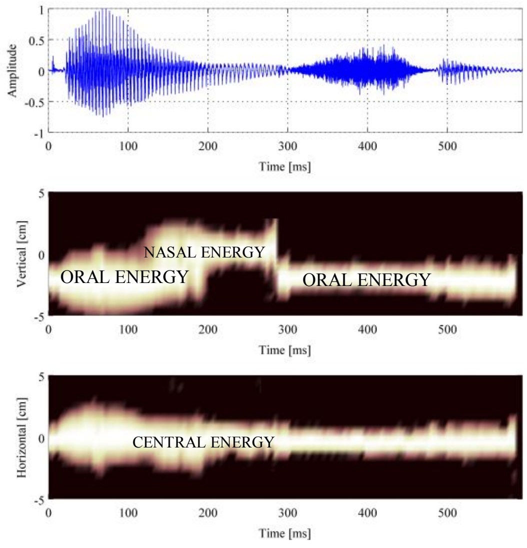

Fig. 7. Spatial distribution of acoustic energy in *pąsy* (Eng. *blushes*).

7. CONCLUSIONS AND FUTURE WORK

The article presents preliminary results of the analysis of the spatial distribution of energy in the acoustic field. The obtained data are very promising for the analysis of sound nasalization and laterality. Still, increasing the resolution of the obtained data would further improve the accuracy of nasality/laterality measurements. The study was performed using the delay-sum algorithm, which is the simplest method of beamforming. In the near future it is planned to examine the possibilities of improving the results with adaptive beamforming algorithms. Other selected elements of analysis will also be changed, including head movement normalization, which was performed manually for the present study. The normalization procedure will be based on tracking the movements of selected fluorescent markers placed on the speakers' faces.

Undoubtedly, the preliminary results of the method developed here and based on the data obtained from the 16-channel microphone array show that it is a non-invasive and effective tool that can be used to objectify the assessment of nasal and lateral articulations.

8. REFERENCES

[1] R. Krakow, and M. Huffman, “Instruments and techniques for investigating nasalization and velopharyngeal function in the laboratory: An introduction.” In: M. Huffman and R. Krakow [eds.] *Phonetics and Phonology: Nasals, Nasalization, and the Velum*, Academic Press, San Diego, pp. 3-59, 1993.

[2] R. J. Baken, and R. F. Orlikoff, *Clinical measurement of speech and voice*, 2nd edition,Taylor and Francis, New York, 2000.

[3] X. Niu, “Measurement, analysis, and detection of nasalization in speech.” *Student Scholar Archive*. Paper 305, 2008.

[4] S.G Fletcher, and M.E. Bishop, “Measurement of nasality with TONAR.” *Cleft Palate Journal*, 7, pp. 610-621, 1970.

[5] R.M. Dalston, D.W. Warren, and E.T. Dalston, “Use of Nasometry as a Diagnostic Tool for Identifying Patients with Velopharyngeal Impairment,” *The Cleft Palate-Craniofacial Journal* 28(2), pp. 184-189, 1991.

[6] F. Bell-Berti, “An electromyographic study of velopharyngeal function in speech,” *Journal of Speech and Hearing Research* 19, pp. 225-240, 1976.

[7] J. Freitas, A. Teixeira, S. Silva, C. Oliveira, M. S. Dias, “Velum Movement Detection based on Surface Electromyography for Speech Interface,” Int. Conf. on Bio-Inspired Systems and Signal Processing (BIOSIGNALS 2014). pp. 13-20, 2014.

[8] M. P. Karnell, E. J. Seaver, and R. M. Dalston, “A comparison of photodetector and endoscopic evaluations of velopharyngeal function,” *Journal of Speech and Hearing Research* 31, pp. 503-510, 1988.

[9] K. L. Moll, and R. G. Daniloff, “Investigation of timing of velar movement during speech,” *Journal of the Acoustical Society of America* 50, pp. 678-684. 1971

[10] A. Serrurier, and P. Badin, “A three-dimensional linear articulatory model of velum based on MRI data,” 6th INTERSPEECH/EUROSPEECH., 2005.

[11] S. Horiguchi, and F. Bell-Bertti, "The Velotrace: a device for monitoring velar position," *Cleft Palate J*, 24(2), pp. 104-11, 1987.

[12] K. Perkell, M. Cohen, M. Svirsky, M. Matthies, I. Garabieta, and M. Jackson, “Electromagnetic midsagittal articulometer (EMMA) system for transducing speech articulatory movements,” *Journal of the Acoustical Society of America* 92, pp. 3078-3096. 1992.

[13] M. Rothenberg, "Measurements of air flow in speech," *J. Speech & Hearing* Res. 20, pp. 155- 176, 1977.

[14] A. Lorenc, “Diagnosis of the pronunciation norm” *Logopedia* 42, pp. 63-87 (http://www.logopedia.umcs.lublin.pl/images/1-278_Logop_42_ANG_ok.pdf).

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

[16] D. Król, “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.

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

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

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

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

[21] D. Król, “Macierze mikrofonowe i głośnikowe”, In: T.P. Zieliński, P. Korohoda and R. Rumian. *Cyfrowe przetwarzanie sygnałów w telekomunikacji: Podstawy, multimedia, transmisja*, PWN, Warszawa, 2014, pp. 665-695.

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