

Magdalena Igras, Stanisław Kacprzak,
Mariusz Maśior, Mariusz Ziółko

Faculty of Computer Science, Electronics and Telecommunication
AGH University of Science and Technology
migras@agh.edu.pl

The acoustic diversity in the phoneme inventories of the world's languages

Abstract. A comparative analysis of multi-language speech samples is conducted using acoustic characteristics of phoneme realisations in spoken languages. Different approaches to investigation of phonemic diversity in the context of language evolution are compared and discussed. We introduced our approach (materials and methods) and preliminary results of research. We built an online database dedicated to voice acquisition and a storage of good quality speech samples collected across the globe. Software designed for automatic extraction and analysis of phonemes was developed and adapted for languages classification. Research involves both experimental and theoretical works that aim at gaining knowledge about phonetic diversity of languages across the world. Additionally, the expected results may be applied to verify the hypothesis of modern languages expansion from Africa, brought to attention by Atkinson.

Keywords: speech processing, languages comparison, phoneme analysis.

1. Introduction

Every language uses only a small part of possible articulation abilities. Young infants have the ability to acquire a spectrum of sounds broader than those existing in any particular language. Their individual articulation abilities are shaped by the culture that motivates them to master only some phonemes.

Archaeological clues may help us determine when the human kind developed the ability of articulation. However, the existence of anatomical potential for uttering of speech sounds does not imply the fact that speech was used. A significant question is when a sequence of phonemes started to convey complex information (Kirby 2007: 669–681). Besides the archaeological or traditional linguistic approaches, genetic methods, computational linguistic and acoustic analyses or their combinations (Cavalli-Sforza *et al.* 1988: 6002–6006) have become widely explored in recent decades.

Phonemic diversity (the number of perceptually distinct units of sound: consonants, vowels and tones in a language) was investigated with a view to estimating when spoken language began (Perreault and Matthew 2012: e35289). Using data from the UCLA Phonological Segment Inventory Database (UPSID) (Maddieson 1984: 162–163), they modelled the phoneme accumulation rate and calculated the time it would take a language to acquire the phonemic diversity observed today in African languages. Perreault and Matthew's study suggests that spoken language appeared early in the history of our species during the Middle Stone Age in Africa, between 150–350 thousand years ago. However, they emphasized that the estimate should be treated as minimal, taking into account the possibility that during evolution the phonemic clock could have pushed back or reset as an effect of the population bottleneck. Information on world languages' phonemes inventory, including their number, types or acoustic properties can be explored not only as an indicator of dating language origin, but also as a source of knowledge on the language evolution process, extinct spoken language reconstruction (Bouchard-Côté 2013: 4224–4229), population ecology and culture studies.

In this paper we present and discuss the state of undertaken research that is meant to bring a new approach to investigate phonemic inventories of languages. Collected data, proposed methodology of acoustic and statistical analyses together with first experiments are described. As a future result of the research, we hope to formulate new scientific evidence on languages' similarities, which may contribute to the discussion on the origin of spoken language.

2. Background: a linguistic founder effect

The term *founder effect* originates from population genetics, and explains the loss of genetic diversity among the migrating parts of a population. In particular, the serial founder effect occurs when populations move over long

distances. Genetic evidence supports the notion that humans originated from Africa, where genetic diversity is the largest and decreases with an increase in distance from Africa (Hay and Bauer 2007: 388–400; Ramachandran *et al.* 2005: 15942–15947; DeGiorgio *et al.* 2009: 16057–16062).

Quentin D. Atkinson in his article “Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa” (2011: 346–349) suggested that the founder effect may operate on human languages as well. Expansion should progressively reduce phonemic diversity with increasing distance from the point of origin, paralleling the serial founder effect observed in evolutionary genetics.

Atkinson examined geographic variation in phoneme inventory size using data taken from 504 languages in the World Atlas of Language Structures (WALS). His statistical analysis confirmed that the number of phonemes in a language is positively correlated with the size of population using these phonemes (Hay and Bauer 2007: 388–400) for both, languages and language family levels. Using the available data together with language location information and supplementing them with a series of linear regression predictors, he estimated a map of likely areas of languages origin and evaluated the model fit with the Bayesian information criterion. The model of the language level points to central and southern Africa as the places where language may come from; also in the family-level analysis, the region of origin is expanded to include the whole of Africa.

In another experiment, Atkinson took into account additional parameters that could have an influence on regional phonemic diversity due to increasing contact within and between groups and creating more opportunities to borrow new phonemes (e.g. the number of languages per unit of area), population densities or language area. Including these factors, he came to the same conclusions.

Atkinson's outcomes are supported by detailed statistical analyses. His results seem to be comparable to those obtained for phenotypic diversity or mitochondrial DNA, but much weaker than for genotypic diversity (Li *et al.* 2008: 1100–1104).

2.1 Atkinson's hypothesis: discussion

Atkinson's article provoked immediate criticism (four technical comments followed by Atkinson's two responses to the comments were published in the *Science Magazine* in 2012) and numerous citations. One of the objections was the simplification of the measurement of the phoneme inventories in the data used for the analysis (Wang *et al.* 2012: 657). Others (Hunley *et al.* 2012:

2281–2288) argue that most of serial founder effect predictions are violated for the phonemic data. According to their opinion, phonemes cannot provide information about more ancient evolutionary processes because they change rapidly. Van Tuyl *et al.* (2012: 657) criticized the method of forced linear regression across all continents and suggested that the approach would be proper within a single continent only. They also questioned the comparability of the phonemic diversity data used by Atkinson. They suggest that taking into account historical processes like migrations, conquests, and borrowings would explain language evolution more credibly than the founder effect solely. Concerns about suboptimal data and biased methodology in Atkinson's work were raised also by Cysuow *et al.* (2012: 657). In Jaeger *et al.* (2013: 1042) technical comments are made and other statistical simulations are performed, and the authors suggested that the type I error rate (the incorrect rejection of the null hypothesis) in Atkinson's analysis is hugely inflated.

Scientific controversy about Atkinson's results has induced intense research in this field motivating us to develop an independent method to assess the hypothesis.

2.2. Motivation and objectives

Inspired by Atkinson's hypothesis of modern languages' expansion from Africa, we want to analyse the acoustic features of phonemes from hundreds of languages from all over the world. We assumed that an analysis of phonemic diversity can be obtained from the analysis of phonemes acoustic features and that this analysis along with the methods used in evolutionary genetics could allow us to create a taxonomy of natural languages.

A set of sounds (allophones) which represent a phoneme in one language can be regarded as more than one phoneme in another language. Thus, analysing a number of phonemes gives us information about languages from a perceptual point of view and an acoustic analysis provides information about languages from a phoneme articulation point of view. The acoustic analysis of phonemes allows us to calculate spectral distances in order to obtain acoustic similarities between a pair of languages.

3. Database

To be able to analyse the acoustic features of languages we need to collect speech samples of hundreds of languages. For that purpose we created a speech sample database (Maşior *et al.* 2013: 79–87; see Fig. 1).

Gathering speech recordings of appropriate quality and length is not an easy task. The results of all of the acoustic analyses can be relevant only if the recording time for each language is sufficiently long (at least several hours). The webpage of Global Recording Network organisation (<http://globalrecordings.net>, DOA: 10 Jan. 2014) is the main source of the collected recordings. The GRN is a provider of Christian evangelistic and discipleship audio-visual materials to the least explored language groups of the world. Moreover, the recordings are collected by the project's webpage (<http://speechsamples.agh.edu.pl>, DOA: 10 Jan. 2014) according to the Web 2.0 paradigm (Fig. 1). We have also obtained data from other speech corpora and extracted speech signals from recordings available on the Internet (online radio and audio books).

Currently, the speech recording database contains recordings for 3563 spoken languages (a total of about 7500 hours). More precise statistics are shown in Tables 1, 2 and 3.

Figure 1. *The research website (<http://speechsamples.agh.edu.pl>) interface*

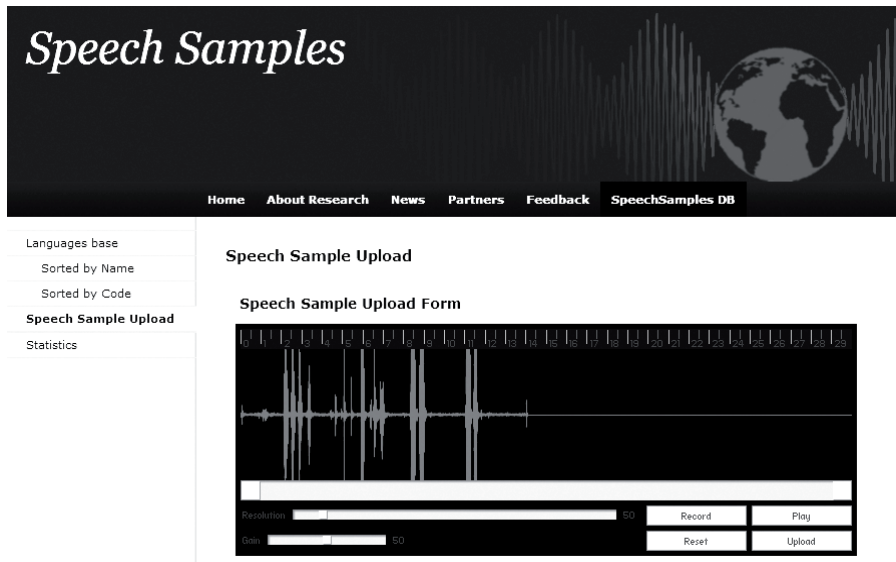


Table 1. *Statistics of speech samples database*

Area	Languages	Duration of recordings [h]
Africa	1241	2618
Asia	1074	2437
Europe	77	452
North America	357	432
Pacific	553	1100
South America	239	423

Table 2. *Length, in hours, of collected speech samples*

Area	Languages recordings length		
	> 1 h	> 5 h	> 10 h
Africa	560	156	37
Asia	558	124	35
Europe	48	16	7
North America	135	12	0
Pacific	288	53	18
South America	114	21	3

Table 3. *Coverage of languages in the database, existing languages number according to Lewis (2009)*

Area	Language population > 20 000		Language population > 500 000	
	Existing languages	In database	Existing languages	In database
Africa	1120	810	199	174
Asia	912	638	249	215
Europe	143	70	60	41
North America	92	69	9	7
Pacific	52	46	0	0
South America	52	35	9	7

4. Methods

Many years of linguistic research have generated precise phoneme classifications for hundreds of languages. The first stage of language comparison analysis was based on linguistic data. Phonemic data (Wang *et al.* 2012) for 574 languages from 6 continents were analysed (Kacprzak *et al.* 2013). Although roughly 9% of world languages were considered, they covered approximately 50% of the world population. The expected numbers of phonemes in the languages on each continent and their standard deviations were calculated. Concerning the languages on each continent, the distribution of vowels and the number of consonants were analysed. Sound diversity among languages differs depending on the continent. All world languages can be divided into four groups. The greatest phonemic diversity was observed (Kacprzak *et al.* 2013) in Africa and Europe, next Asia, followed by North America. Oceania and South America, which belong to the last group, were marked by relatively low phonemic diversity.

Nowadays, all these results and conclusions can be verified by a computer analysis of speech recordings. We should expect different dependencies between speech signals for world languages. Such a computer analysis uses the signal processing method and is based on the acoustic properties of speech.

To provide the acoustic analysis we need to segment speech signals into acoustically uniform pieces that correspond to phonemes. Speech segmentation is a core process in the automatic analysis of phonemes.

The collected recordings are segmented into elementary units (Ziółko *et al.* 2010: 2234–2237) using spectral methods based on the discrete wavelet transform (Daubechies 1992). The methodology for speech recognition developed during the past years provides accurate segmentation and usable implementations. Each fraction of the speech signal is split by digital, low-pass and high-pass filters. Low frequencies have narrow bandwidths and are investigated with a finer resolution, while high frequencies have wide bandwidths, which results in a lower resolution. It is possible to analyse speech in frequency ranges corresponding to a perceptual scale by applying wavelet packet transform (a generalization of wavelet transform), in which low-pass and high-pass bands are split additionally (see Fig. 2).

Figure 2. Frequency sub-bands obtained from wavelet packet transform with a perceptual scale (Ziółko et al. 2010: 2234–2237)

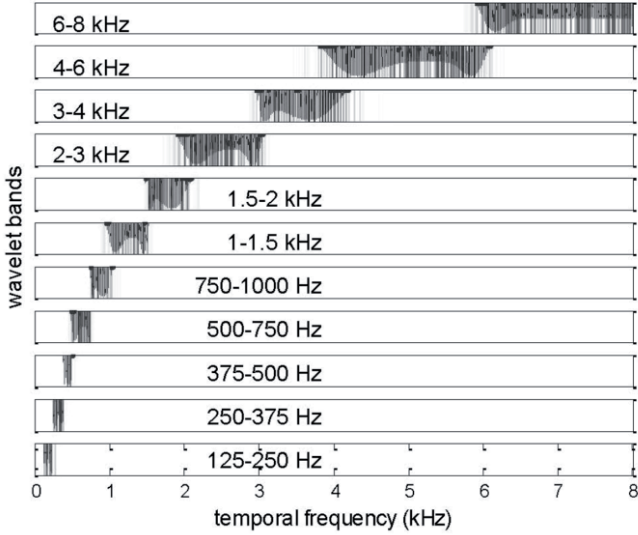
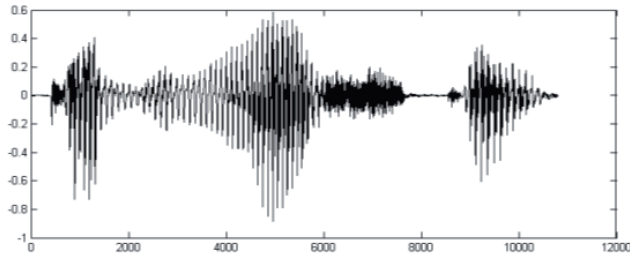
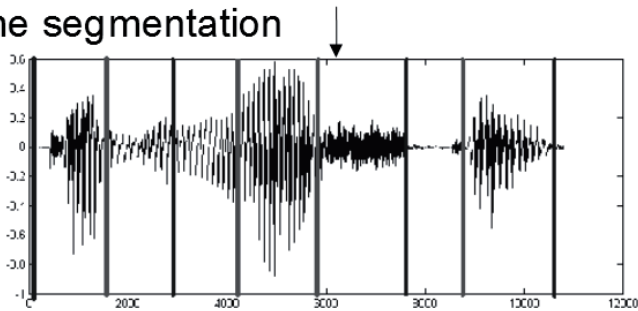


Figure 3. Results of Speech signal segmentation



Phoneme segmentation



The phoneme segmentation (Fig. 3.) and analysis algorithm is based on the analysis of fractions of speech energy in each frequency band. Phonemes can be described as quasi-stationary processes. The role of the segmentation algorithm is to detect significant transitions of energy among frequency sub-bands (Gałka and Ziólko 2010: 128–135). Then, qualitative and quantitative analyses of the extracted phoneme features are performed.

Parameterization of phonemes is obtained as the mean energy of acoustic signal in each of twelve frequency bands, weighted by Hamming window to avoid the boundary interactions. In that way each phoneme is represented as a twelve dimensional vector.

5. Evaluation and initial results

The first verification of inference possibilities from the results of the fully automatic analysis of speech signals was based on the generation of distances matrix from phoneme clouds, calculated for several dozens of languages. Languages were modelled by a single Gaussian distribution, obtained from parameterised phonemes realisations from the gathered recordings.

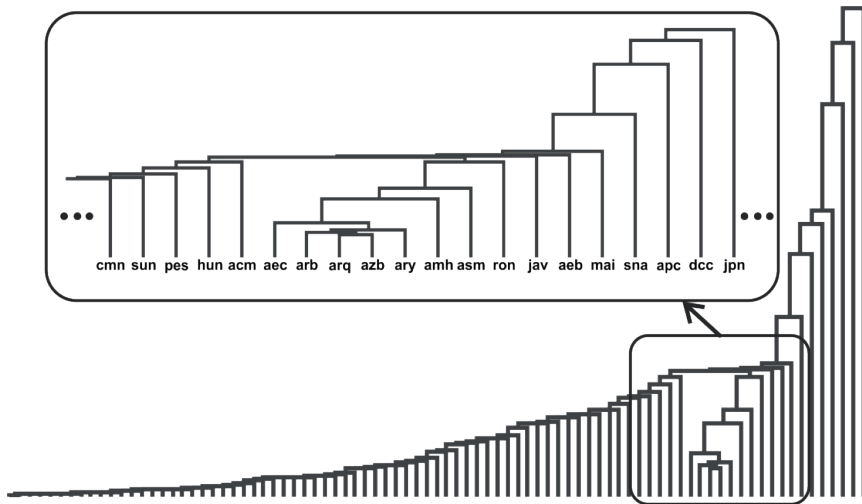
The top 85 languages (by recording length) were analysed. The Bhattacharyya distance (Brian and Barnard 1996: 2005–2008) was used for calculating similarities between the Gaussian distributions. The UPGMA (Murtagh 1984: 101–113) method resulted in a dendrogram, which is presented in Fig. 4.

Although all of the clustered results are not very encouraging, a separate branch with mostly Arabic languages (ISO 639-1 codes) can clearly be seen:

- aec – Arabic, Sa'idi Spoken,
- arb – Arabic, Standard,
- arq – Arabic, Algerian Spoken,
- azb – Azerbaijani, South,
- ary – Arabic, Moroccan Spoken.

The preliminary results of such approach are consistent with linguistic knowledge (Lewis 2009). Most of the other languages belong to one big cluster because the distances between them are similar. We expect that it will be possible to obtain a better language division and hierarchisation using a more precise method of language modelling (Gaussian Mixture Models for language phonemes).

Figure 4. The UPGMA dendrogram of language classification (ISO 639-1 codes as language designation)



6. Conclusion

In this article we present the motivation for our ongoing research on phonemes of world languages classification and we explain the core methodology of signal processing used in the research. The precise acoustic analysis of multi-linguistic speech may provide answers to the following question: which phonemes are used in different parts of the world and what are their individual features. The first results of the acoustic comparison between languages show that the adopted approach is justified and might result in interesting conclusions.

Acknowledgements

The project was funded by the National Science Centre allocated on the basis of a decision DEC-2011/03/B/ST7/00442.

References

Atkinson, Q. D. (2011). Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa. *Science* 332.6027: 346–349.

- Atkinson, Q. D. (2012). Response to Comments on: Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa. *Science* 335.6069: 657.
- Atkinson, Q. D. (2012). Response to Comment on: Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa. *Science*: 1042.
- Bouchard-Côté, A., D. Hall, T. L. Griffiths, and D. Klein. (2013). Automated reconstruction of ancient languages using probabilistic models of sound change. *PNAS* 110.11: 4224–4229.
- Brian, M. and E. Barnard. (1996). Phone Clustering Using the Bhattacharyya Distance. *Proceedings of Fourth International Conference on Spoken Language* 4: 2005–2008.
- Cavalli-Sforza, L. L., A. Piazza, P. Menozzi, and J. L. Mountain. (1988). Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *PNAS* 85.16: 6002–6006.
- Cysouw, M., D. Dediu, and S. Moran. (2012). Comment on: Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa. *Science* 335.6069: 657.
- Daubechies, I. (1992). *Ten Lectures on Wavelets*. Philadelphia: Society for Industrial and Applied Mathematics.
- DeGiorgio, M., M. Jakobsson, N. A. Rosenberg. (2009). Explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa. *PNAS* 106.38: 16057–16062.
- Gałka, J. and M. Ziółko. (2010). Wavelet Speech Feature Extraction Using Mean Best Basis Algorithm. *Advances in Nonlinear Speech Processing. Lecture Notes in Computer Science* 5933: 128–135.
- Hay, J., and L. Bauer. (2007). Phoneme inventory size and population size. *Language* 83.2: 388–400.
- Hunley, K., C. Bower, M. Healy. (2012). Rejection of a serial founder effects model of genetic and linguistic coevolution. *Proceedings of the Royal Society B: Biological Sciences* 279.1736: 2281–2288.
- Jaeger, T. F., D. Pontillo, and P. Graff. (2012). Comment on: Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa. *Science* 335.6072: 1042.
- Kacprzak, S., M. Ziółko, M. Maşior, M. Igras, K. Ruszkiewicz. (2013). Statistical Analysis of Phonemic Diversity in Languages Across the World. *Proceedings of XIX National Conference on Applied Mathematics in Biology and Medicine*.
- Kirby, S. (2007). The evolution of language. In R. Dunbar and L. Barrett (eds.), *Oxford Handbook of Evolutionary Psychology*. 669–681.
- Lewis, P. M. (2009). *Ethnologue, Languages of the World*. 16th Edition. Dallas: SIL International.
- Li, J. Z., D. M. Absher, H. Tang, A. M. Southwick, A. M. Casto, S. Ramachandran, H. M. Cann, G. S. Barsh, M. Feldman, L. L. Cavalli-Sforza and R. M. Myers.

- (2008). Worldwide Human Relationships Inferred from Genome-Wide Patterns of Variation. *Science* 319.5866: 1100–1104.
- Maddieson, I., and S. F. Disner. (1984). *Patterns of sounds*. New York: Cambridge University Press. 162–163.
- Mąsior, M., M. Igras, M. Ziółko, S. Kacprzak. (2013). Database of speech recordings for comparative analysis of multi-language phonemes. *Studia Informatica* 34.2B: 79–87.
- Murtagh, F. (1984). Complexities of Hierarchic Clustering Algorithms: the State of the Art. *Computational Statistics Quarterly* 1: 101–113.
- Perreault, C. and S. Mathew. (2012). Dating the Origin of Language Using Phonemic Diversity. *PLoS ONE* 7.4: e35289.
- Ramachandran, S., O. Deshpande, C. C. Roseman, N. A. Rosenberg, M. W. Feldman, and L. L. Cavalli-Sforza. (2005). Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *PNAS* 102.44: 15942–15947.
- Van Tuyl, R. and A. Pereltsvaig. (2012). Comment on: Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa. *Science*: 657.
- Wang, C., Q. Ding, H. Tao and H. Li. (2012). Comment on: Phonemic Diversity Supports a Serial Founder Effect Model of Language Expansion from Africa. *Science* 335.6069: 657.
- Ziółko, M., J. Gałka, B. Ziółko and T. Drwięga. (2010). Perceptual Wavelet Decomposition for Speech Segmentation. *Proceedings of the Interspeech*: 2234–2237.