Distances between phonemes?

It is not a paper in a scientific meaning – it is just a conspect for educational purposes, in which I wrote something I read as far as I understood it. To write a real paper, a person should have something to say to the world, but to get a such level in some tematics, it is necessary to spent a lot of time, what is impossible in two months, at least if you have full time job, family etc.

I think, it is in the nature of people to compare different objects and to try to find how fare they are one from other. Some times it is usefull, some times – not, but it is allways interesting.

Languages, of course, are not an exception. The interestingness is confessed, but there is some usefullness too. As usual for computational linguistics' fields, we can divide this problem in two: for speech and for texts. (Of course, texts should be in phonetic transcription or we should have a procedure how to convert them to it.) The main difference is that in the case of speech we talk about a real sound, but in the case of phonetic transcription – some kind of ideal middle-values of a tongue. Also the fields of usage are different: the distortion measures used on phonogramms are used for speech recognition, but distances between fixed phonems – for theoretical field - language comparison, particularly – dialectometry.

The first one better fits to the subject of the course, however I personally was more interested in the second one. I am very interested in dialectometry of Baltic and Slavic languages, and I tried to find a stuff for it, of course. The main idea was to find something about measures between sounds of a language or a group of some languages. Unfortunately, I did not found exactly what I wanted. Of course, people who uses the Levenshtein's algorithm for phonetic strings comparison, understand that a binary substitution cost between phonetic characters (equivalent – not equivalent) is not a good way to do it. However, most authors try to find an easy intuitive solution, because a well done theoretically grounded solution could be too time-capacious. For example, in [1] "the subsitution of one letter for another involving only a change in diacritic was valued at 0.2 (rather than 2)". In [2] the author choosed a more complex solution, wich is still just a darn: he "distinguished them on the basis of twelve phonetic features that systematically account for all of the distinctions in Wagner's inventory: nasality, stricture, laterality, articulator, glottis, place, palatalization, rounding, length, height, strength, and syllabicity. The features were given discrete ordinal values scaled between 0 and 1, the exact values being arbitrary. For example, *place* took on the values *glottal=O, uvular=0.1, postvelar=0.2, velar=0.3, prevelar=O.~, palatal=0.5, alveolar=0.7, dental=0.8, and labial=l.* The distance between any two phones was judged to be the difference between the feature values, averaged across all twelve features. These distances were used instead of uniform 1-unit costs in computing Levenshtein distance. The resulting metric was called *feature string comparison.*" It also says that he "know of no comprehensive study on the differences between phones, at least not for all 277 contrasts made by Wagner". That's why he choosed the weighting described above.

In [3] author analyses this question and writes about different ways how to do it, and it would be interesting for the reader too: "Ladefoged (1975) devised a phonetically-based multivalued feature system. This system was adapted by Connolly (1997) and implemented by Somers (1998). It contains about twenty articulatory features, some of which, such as *Place*, can take as many as ten different values, while others, such as *Nasal*, are basically binary oppositions. For example, the feature *Voice* has five possible values: [glottal stop], [laryngealized], [voice], [murmur], and [voiceless]. Feature values are mapped to numerical values in the [0,1] range. The main problem with both Somers's and Connolly's approaches is that they do not differentiate the weights, or *saliences*, that express the relative importance of individual features. For example, they assign the same salience to the feature *Place* as to the feature *Aspiration*, which results in a smaller distance between [p] and [k] than between [p]

and [ph]. In my opinion, in order to avoid such incongruous outcomes, the salience values need to be carefully differentiated; specifically, the features *Place* and *Manner* should be assigned significantly higher saliences than other features. Although there is no doubt that not all features are equally important in classifying sounds, the question of how to how to assign salience weights to features in a principled manner is still open. Nerbonne and Heeringa (1997) experimented with weighting each feature by information gain but found that it actually had a detrimental effect on the quality of alignments. Kessler (1995) mentions the uniform weighting of features as one of possible reasons for the poor performance of his feature-based similarity measure. Covington (1996) envisages "using multivariate statistical techniques and a set of known 'good' alignments" for calculating the relative importance of each feature, but provides no specific details. In my opinion, it seems feasible to derive the saliences automatically from a large corpus of aligned cognates by adapting methods developed for molecular biology (Durbin et al., 1998). Unfortunately, such a representative training set is not readily available because the task of establishing the correct alignment of cognates by hand is very time-consuming. Moreover, any selection of the training data would bias the similarity function towards particular languages. An important advantage of the feature-based metrics is a small number of parameters. It would be ideal to have, as stated by Kessler (1995) in his computational analysis of Irish dialects, "data telling how likely it is for one phone to turn into the other in the course of normal language change." Such universal scoring schemes exist in molecular biology under the name of Dayhoff's matrices for amino acids (Dayhoff et al., 1983). However, the amount of data available in dialectology is many orders of magnitude smaller than what has already been collected in genetics. Moreover, the number of possible sounds is greater than the number of amino acids. The International Phonetic Alphabet, which is a standard for representing phonetic data, contains over 80 symbols, most of which can be modified by various diacritics. Assembling a substitution matrix of such size by deriving each individual element is not practicable. In the absence of a universal scoring scheme for pairs of phonetic segments, the calculation of similarity scores on the basis of articulatory phonetic features with salience coefficients is a good working solution."

However, also the physical-accoustic side of the topic was interesting to me – it's like a try to approach to the problem from the other side. Here all is much concrete, because there are real data it is possible to work with – speech signals – and it is not necessary to find such hard theoretical things to make it usable like with phonetically transcribed texts.

The main idea is decribed very well in [7]: "A distortion measure is an assignment of a nonnegative number to an input/output pair of a system. The distortion between an input or original and an output or reproduction represents the cost or distortion resulting when that input is reproduced by that output." Technically it looks like that [7]: "All of the speech distortion measures considered here depend on their sampled speech waveforms only through their second-order properties-their sample autocorrelations or spectral models. These distortion measures are most easily defined in the spectral domain, though their evaluation is most often carried out without reference to that domain. ... A spectral distortion measure is a function of two spectral densities, f and g for example, which assigns a nonnegative number $d(f,g)$ to represent the distortion in using g to represent f. The most common of such measures are difference distorton measures where one uses an L_p, norm on the difference $f - g$."

In the [15] author describe and compaire many different distortion measures: Itakura-Saito, log likelihood ratio (Itakura), weighted likelihood ratio, weighted slope metric and cepstrum.

In [10] is given a well done classification of distortion measures. They are classified in three general types: generalized kolmogorov variational distance, f-divergence and Chernoff distance. Author prove, that other popular distortion measures are just special cases of those.

1. Nerbonne, J., Heeringa, W., van den Hout, E., van der Kooi, P., Otten, S., van de Vis, W. *Phonetic Distance between Dutch Dialects*.
2. Kessler, B. *Computational dialectology in Irish Gaelic*.
3. Kondrak, G. *Phonetic alignment and similarity*. 2003.
4. Caballero, M., Moreno, A., Nogueiras A. *Data Driven Multidialectal Phone Set for Spanish Dialects*.
5. Kirchhoff, K. *Syllable-level desynchronisation of phonetic features for speech recognition*.
6. Juola, P., Zimmermann, P. *Whole-Word Phonetic Distances and the PGPfone Alphabet*.
7. Gray, R., Buzo, A., Gray, A., Matsuyama, Y. *Distortion Measures for Speech Processing*.
8. Mak, B., Barnard, E. *Phone Clustering Using the Bhattacharyya Distance*.
9. Leusch, G., Ueffing, N., Ney, H. *A Novel String-to-String Distance Measure With Applications to Machine Translation Evaluation*.
10. Lee, Y. *Information-Theoretic Distortion Measures for Speech Recognition*. 1991.
11. Klatt, D. *Prediction of Perceived Phonetic Distance from Critical band Spectra: a First Step*.
12. Beresfor-Smith, B., Breckling, J., Schroder, H. *Systolic Devices for Speech Processing*.
13. Taylor, P., Black, A. *Speech Synthesis by Phonological Structure Matching*. 1999.
14. Eaton, T. *Exploring the Sub-phonemic Level within the Word Recognition System.*
15. Nocerino, N., Soong, F., Rabiner, L., Klatt, D. *Comparative Study of Several Distortion Measures for Speech Recognition.*
16. Аграновский А.В., Леднов Д.А., Репалов С.А., Телеснин Б.А. *Система автоматической классификации фонем русского языка при её обучении методом группового учёта аргументов.*