Formants and Nasalisation in the Detection of URI from Telephone-Bandwidth Speech

Ladan Baghai-Ravary
Advanced Speech Processing Team Leader, Aculab PLC
ladan.ravary@aculab.com

Most research into the effects of Upper Respiratory Infection (URI) on voice and speech, has concentrated on aspects of the voice source and its stability (jitter, shimmer, harmonic-to-noise ratio, etc.). These features can indeed be indicative of URI [Lopes, Cavalcante and da Costa, 2014], especially when applied to sustained phonations or other tightly controlled utterances (containing particular phonetic contrasts, for example).

However, there are still doubts about the use of "perturbation" parameters outside the laboratory [Hillenbrand, 2011], at least in part because of the variability observed between different measures of jitter, shimmer, etc. in different publications and different environments. This is quite apart from the reported difficulty in obtaining useful values for these measures from natural speech, as opposed to sustained phonation.

In this paper we take two approaches to extend the scope of such studies. Firstly we estimate perturbation parameters without explicit identification of pitch epochs, improving robustness and facilitating the analysis of uncontrolled recordings, and in particular, of speech over the telephone. Secondly we introduce additional measures, to characterise properties of vocal tract resonances, including nasalisation, and the dynamics of continuous speech. Some of these characterise a complete utterance, while others operate on a single unbroken interval of voiced speech, to capture information at a prosodic-level.

To test the relative merits of the various measures, we analysed a large portion of the URTIC corpus [Cummins et al., 2017]. We then calculated the Kruskal-Wallis H-test [Kruskal and Wallis, 1952] between the observed distributions of each measure, for healthy and URI speech samples respectively. We used this method in preference to ANOVA analysis because some of the observed distributions were clearly non-Gaussian; since the H-test is based purely on the rankings of values, rather than the values themselves, it allows a comparison to be made between non-Gaussian distributions.

The results of this analysis indicate that the most discriminative features for detection of URI from telephone-bandwidth speech are those related to the resonances and anti-resonances of the vocal tract. However, both short-term changes in signal amplitude (shimmer), and longer-term variations, measured on a prosodic timescale, are also significant.

References

Cummins, N, et al., "You sound ill, take the day off: Classification of speech affected by Upper Respiratory Tract Infection" to appear in Proc. EMBC 2017, Jeju Island, South Korea (2017).

Hillenbrand, JM; "Acoustic Analysis of Voice: A Tutorial" in Perspectives on Speech Science and Orofacial Disorders, 21:31-43 (2011).

Kruskal, WH, and Wallis, WW; "Use of Ranks in One-Criterion Variance Analysis", Journal of the American Statistical Association, Vol. 47, Issue 260, pp. 583-621 (1952).

Lopes, LW, Cavalcante, DP, and da Costa, PO; "Severity of Voice Disorders: Integration of Perceptual and Acoustic Data in Dysphonic Patients" in CoDAS 2014;26(5):382-8 (2014).