spec\_0001-00-00
There is no discontinuity in the *mfcc* version, apart from the fact that there is an audible [m] rather than [n]. The *formants\_ABS* version is best, there is a clear [n] and no discontinuity. The *formants\_EUCL* has continuous formants but differences in spectral energy cause a strong discontinuity.

sent\_0619-00-00
The *mfcc* version is clearly superior, though there are some temporal issues (local articulation rate is unnatural in *zbylé svědky*). The *formants\_ABS* versions contains a lot of problems: unnatural F0 jumps (not discontinuities within syllables but entire syllables), spectral distribution discontinuity in [ou] in *vyslechnou*, *svědky* sounds like [svjɪtkɪ].

sent\_0953-00-00
A very interesting case. The *mfcc* version is not perfect but the overall quality really seems the best. The *formants\_ABS* version is, in fact, better in all the traditional respects except for one detail which Czech listeners seem to be quite sensitive to – the dental realization of [d] in *dům*. In Czech, [t] is a denti-alveolar sound (more front), while [d] is alveolar to post-alveolar (more back), and a denti-alveolar pronunciation of [d] is regarded as a speech defect (*cf.* the two versions from this perspective). The *formants\_EUCL* version contains quite a lot of discontinuities: F0 in *školu* and *dům*, and spectral energy in *dům*.

sent\_3170-00-00
Again, the *mfcc* version is not without small problems, but the *formants\_ABS* version has serious problems in the last word, *odložil*.

sent\_4721-00-00
The *mfcc* version has quite a lot of problems: the [uː] in *problémů*, F0 jumps in *nebylo*. But these are completely overshadowed by the horrible timing in the *formants\_EUCL* version, where the first two syllables of *problémů* are very short, the extremely long [o] in *nebylo*, as well as some smaller F0 jumps.

To conclude, formant frequencies are only one aspect in the spectral domain in which a discontinuity may arise. Since the bandwidths were not considered – and their computation, at least in Praat, does not seem to be very reliable – the complex of spectral information is quite limited. The distribution of energy in the spectrum (which can be expressed as spectral slope) is not accounted for, and differences in that leads to audible discontinuities. MFCC coefficients, on the other hand, should account for the frequencies of the main spectral components (formants), as well as spectral energy distribution – albeit in a way which is not transparent or interpretable.

It would therefore be interesting to compare two versions of synthetic speech where one version is based on MFCCs alone, and the other version combines formant frequencies along with some expression of spectral slope.