BAAP2010
COLLOQUIUM
LONDON

Reliability of Formant Measurements from Lossy Compressed Audio

James Bulgin*, Paul De Decker* and Jennifer Nycz**
*Memorial University of Newfoundland / **University of York

Background

Widely-used compressed audio formats such as MP3 achieve their small file-size through lossy compression. **Compressed audio** simplifies the logistics of managing large audio corpora and facilitates online sharing of such corpora. Moreover, some sociolinguistically interesting recordings may only be available in compressed format.

Audio compression is also used in popular VoIP services such as Skype and Google Video Chat. Services such as these are potentially useful for collecting linguistic data from speakers who might not otherwise be accessible to the researcher, while saving substantial travel costs and time.

However, compression destructively modifies the audio, simplifying some frequency content and discarding others.

Before compressed audio can be used for phonetic research, it is therefore crucial to answer the following question: **To what degree does compression affect the reliability of acoustic analysis? (**Gonzalez et al. 2001, 2003).

Methods

Sociolinguistic interviews were recorded with four speakers (2 male, 2 female) to 24-bit 44Khz wav files. Copies of each interview were then compressed to a constant bitrate MP3 at 320 kbps, 128 kbps, (16-bit, 44KHz) and 64 kbps (16-bit, 22KHz): low, moderate, and high compression, respectively.

Two of the interview recordings (Speaker D and Speaker O, 1 female and 1 male) were also transmitted over Skype and recorded on another computer. The audio was piped directly from the source computer's playback into Skype (there were no additional microphones used), so no additional sources of noise were introduced other than the effects of Skype's audio compression algorithm.

Tokens were chosen from each interview to establish the complete vowel space for each speaker. An average of 5 tokens was selected for each of 11 vowels. F0 through F4 were measured for each token at the same timestamp in each file format, and these measurements were compared across file formats for each speaker using repeated measures ANOVA.

Type (file formats) and Vowel were included as independent variables, and an interaction term was added to each model to see whether Type varied in its effects across different Vowels.

Discussion

Main Findings:

1. **We found no loss in the reliability of measurements of F0 in any compressed format** (plots & statistical analyses not shown here; there was no effect of compression Type on F0)

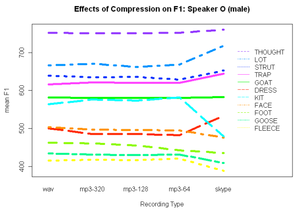
2. **The measurements from MP3 compression did not significantly differ from those taken from the WAV format**, though there is some evidence of degradation with the lossiest form (MP3-64), especially in F3 and F4.

3. The compression algorithm behind Skype transmissions significantly altered formants which are relevant to vowel identification. **Speaker O shows a particularly clear Skype-effect: his vowel space is stretched along both dimensions**, such that high vowels are made measurably “higher” (lowered F1) and low vowels are made “lower” (higher F1), while back vowels are more “backed” (lowered F2) and front vowels are more “fronted” (higher F2). Speaker D does not show the same kind of distortion, as her vowel space is stretched at the high front corner, but compressed elsewhere; however, her data also reveal an interaction between compression Type and Vowel.

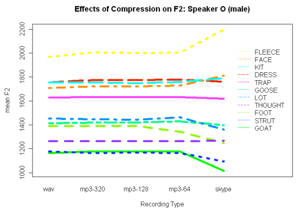
Based on these findings, it seems that **speech data stored in MP3 format may be suitable for phonetic analysis, as there is little if any distortion of the most linguistically relevant formant information.**
However, speech data obtained via Skype should be used with caution if at all, as the format distorts the vowel space in nonlinear ways.

Results

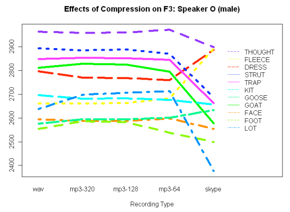
Effects of compression types on different vowels



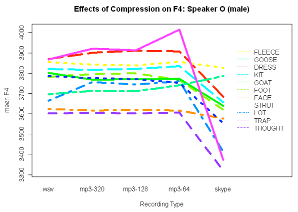
Main effect of Vowel on F1 (F(10, 44)= 14.234, p<0.001) Interaction between Vowel & Type (F(40, 176)= 2.9476, p<0.001). Low vowels have higher F1s in Skype recordings, while high vowels have lower F1s.



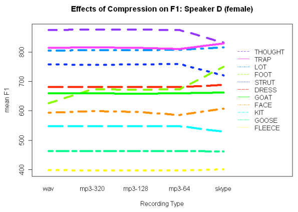
Main effect of Vowel on F2 (F(10, 44)= 7.734, p<0.001) Interaction between Vowel & Type (F(40, 176)= 3.483, p<0.001). Front vowels have higher F2s in Skype recordings, while back vowels have lower F2s.



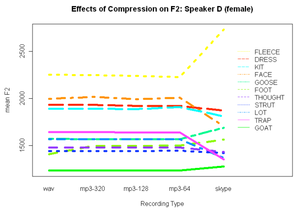
No significant main effects of Vowel or Type on F3, nor interactions, yet Skype does appear to alter F3 for different vowels in different ways.



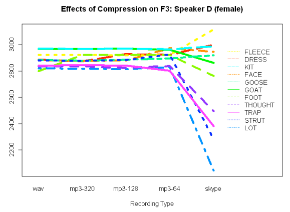
Main effects of Type on F4 (F(4,176)=9.136, p<.001), with Skype appearing to lower F4 across vowels.



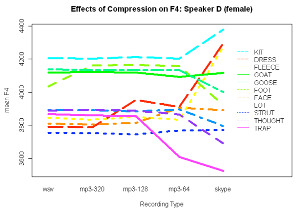
Main effect of Vowel on F1 (F(10, 43)= 14.537, p<0.001). No significant effect of Type or interactions.



Main effect of Vowel on F2 (F(10, 43)= 3.768, p=0.001) Interaction between Vowel & Type (F(40, 172)= 2.993, p<0.001). The F2 of FLEECE is dramatically increased in the Skype recording.

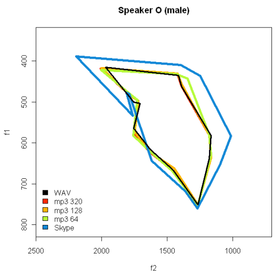


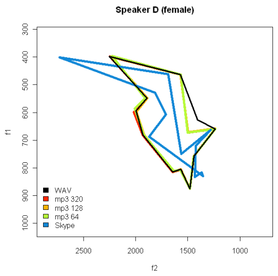
Main effect of Type on F3 (4,172)= 6,869, p<0.001) Interaction between Vowel & Type (F(40, 172)= 1.986, p=0.001). Skype has an overall lowering effect on F3, though this affects certain vowels more than others.

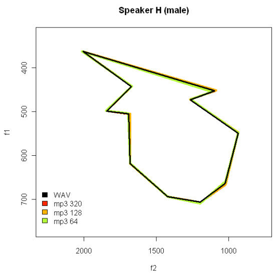


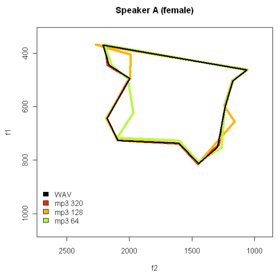
No significant main effects of Vowel or Type on 4, nor interactions, though much distortion in Skype format.

Effects of compression types on vowel space


Speaker O (male)
f1
f2
WAV
mp3 320
mp3 128
mp3 64
Skype


Speaker D (female)
f1
f2
WAV
mp3 320
mp3 128
mp3 64
Skype


Speaker H (male)
f1
f2
WAV
mp3 320
mp3 128
mp3 64


Speaker A (female)
f1
f2
WAV
mp3 320
mp3 128
mp3 64