

Chapter III-16 — Text Encodings

Starting with Igor Pro 9, during experiment loading, if text is valid as UTF-8, including all ASCII text, it is loaded as UTF-8 regardless of the text encoding stored in the experiment file. The purpose is to transition away from obsolete text encodings.

Also starting with Igor Pro 9, if a plain text file has a UTF-8 byte order mark it is loaded as UTF-8 even if it contains invalid byte sequences. Invalid bytes are displayed using the Unicode replacement character (U+FFFD).

If Igor can not otherwise find a valid text source encoding it displays the Choose Text Encoding dialog. This dialog asks you to choose the text encoding to use for the plain text file being opened.

Override Experiment Checked

If the Misc→Default Text Encoding→Override Experiment menu item is checked then the order changes to this:

```
Byte order mark
TextEncoding pragma if present (for procedure files only)
The default text encoding
The specified text encoding (described below)
UTF-8 if the text contains non-ASCII characters
Choose Text Encoding dialog (described below)
```

In this mode the default text encoding, as set in the Misc→Default Text Encoding submenu, is given a higher priority. This should be used only if the normal mode fails to give the correct results. If you turn Override Experiment on, make sure to select the appropriate default text encoding from the submenu. You should turn Override Experiment on only in emergencies and turn it off for normal use.

Starting with Igor Pro 9, during experiment loading, if text is valid as UTF-8, including all ASCII text, it is loaded as UTF-8 regardless of the Override Experiment setting. The purpose is to transition away from obsolete text encodings.

Potential Problems Determining Plain Text File Text Encodings

The rules listed above have the following potential problems:

TextEncoding pragma (for procedure files only)

The pragma could be wrong. For example, if you set TextEncoding="MacRoman" and then change the file to UTF-8 in an external editor, it will still succeed in Igor but give the wrong characters. This is a programmer error and it is up to you to fix it.

The specified text encoding

There is no specified text encoding for plain text files in a pre-Igor6.3x experiment.

In Igor6.3x and later, Igor includes text encoding information in the experiment restart procedures for each plain text file to be opened. This information is the "specified text encoding".

In Igor6.3x the specified text encoding can be wrong. For example, if the user adds some Japanese characters to a graph annotation, the recreation macro will include Shift-JIS. If the font controlling the procedure window is MacRoman or Windows, the specified text encoding will be MacRoman or Windows and the Japanese characters will show up wrong in Igor7 or later.

UTF-8 if the text contains non-ASCII characters

MacRoman, Windows and ShiftJIS can all masquerade as UTF-8 though this will be rare. By "masquerade" we mean that the non-ASCII bytes in the text are legal as UTF-8 but the text really is encoded using another text encoding.

For a procedure file you can fix this by adding a TextEncoding pragma.