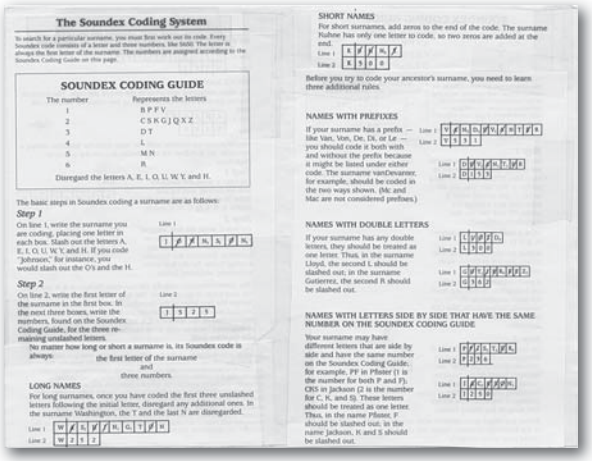
The Soundex Coding System

SOUNDEX CODING GUIDE

The number	Represents the letters
1	B P F V
2	C S K G J Q X Z
3	D T
4	L
5	M N
6	R

Disregard the letters A, E, I, O, U, W, Y, and H.

Step 1

Step 2

LONG NAMES

SHORT NAMES

NAMES WITH PREFIXES

NAMES WITH DOUBLE LETTERS

NAMES WITH LETTERS SIDE BY SIDE THAT HAVE THE SAME NUMBER ON THE SOUNDEX CODING GUIDE

Thinking in Soundex

By Claire Prechtel-Kluskens

Online genealogy websites have made it absurdly easy to locate our ancestors in census and other original historic records. Type in the name, and, voilà, there's the ancestor. But it's not always that easy. Sometimes researchers "can't find" an ancestor. Why? One common reason, but not the only one, is that the ancestor's name was spelled in the original historic records differently than the way we expect. The census enumerator or other official often wrote the name according to the "ear" test—what did he think was said?[1] This results in a plethora of spellings, such as foreign names like Johanson being written as Johnson, or twenty-one different spellings of McClendon.[2] To deal with these problems, we need a strategy. "Thinking in Soundex" is a useful tool to understand and develop potential alternative spellings for a surname.

"Soundex" is a phonetically based indexing system developed and patented by Robert Russell, who then licensed the system rights to Remington Rand Company. That company then trademarked the name "Soundex" and promoted it as a tool to increase government and corporate efficiency.[3] In 1936, the Bureau of the Census chose the Soundex system as the means of organizing its new Works Progress Administration (WPA)-created card index for the 1900 census. Subsequently, the Soundex system was used for the WPA-created indexes to the 1880 and 1920 census, the Census Bureau's 1910 Miracode, WPA-created indexes to records of the Immigration and Naturalization Service, and by state governments and other organizations. Today, some online genealogy sites allow Soundex.

Soundex is based on phonics: similar sounding consonants are coded the same way so that similar sounding surnames get the same Soundex code. However, there are variations in spelling that affect both the Soundex code as well as a name's place in the alphabetical sequence. It is in these situations that knowing and applying the code helps force us to consider wildly different spellings. The nuts and bolts of Soundexing are given in the sidebar, so let's focus on the creative ways it can help us understand surname variations.

Vowel substitution

Vowels are not coded in Soundex, so substituting one vowel for another doesn't change the Soundex code. Smith, coded as S-530, thus has many potential equivalents, such as Smythe, Smath, Smeth, and Smoth. Clancy (C-452) is equivalent to Clancie, and Reid (R-300) is equivalent to Reed, Read, and Reade.

Initial consonant substitution

Both straight alphabetical indexes and Soundex are based on looking for a surname at its expected place in the alphabet. Clancy (C-452) is therefore expected to be among surnames beginning with "C." If the name was written in the original record as Klancy (K-452), or Glancy (G-452), we won't find it, because of our assumption that the name begins with C. Thus, the principle of initial consonant substitution teaches us to realize that if the name starts with C, for example, then we should consider its Soundex equivalents–G, J, K, Q, S, X, and Z– as *possible* alternative beginning letters to the name. Other examples are Prechtel (P-623) recorded as Brechtel (B-623), Siml (S-540) as Chimel (C-540), Cerwin (C-650) as Servin (S-615), and Zavodny (Z-135) as Cawodny (C-150) and Savonia (S-150).[4]

Initial consonant dropping

This provides similar problems to initial consonant substitution. A name like Clancy (C-452) might instead be recorded as Lancy (L-520), Harbaugh (H-612) as Arbaugh (A-612), or James McDonald (M-235) might instead be recorded James M. Donald (D-543), so our searches need to take those potential variations into consideration.

Middle consonant substitution

Middle consonant substitution can make a difference. Clancy and Clamsy (both C-452), Madison and Matteson (both M-325), and Stevenson and Stephenson (both S-315) are

Soundex coding guide

1 B, F, P, V
2 C, G, J, K, Q, S, X, Z
3 D, T
4 L
5 M, N
6 R

Every Soundex code consists of a letter and three numbers, such as W-252 for Washington, or L-000 for Lee. Zeroes are added at the end if necessary; extra consonants are disregarded.

Special rules

1. Names with double letters. If the surname has any double letters, they should be treated as one letter. For example: Gutierrez is coded G-362 (G, 3 for the T, 6 for the first R, second R ignored, 2 for the Z).
2. Names with letters side-by-side that have the same Soundex code numbers. If the surname has different letters side-by-side that have the same number in the Soundex coding guide, they should be treated as one letter. Examples:

   Pfister is coded as P-236 (P, F ignored, 2 for the S, 3 for the T, 6 for the R).

   Jackson is coded as J-250 (J, 2 for the C, K ignored, S ignored, 5 for the N, 0 added).
3. Names with prefixes. If a surname has a prefix, such as Van, Con, De, Di, La, or Le, code both with and without the prefix because the surname might be listed under either code. Note, however, that Mc and Mac are not considered prefixes. For example, VanDeusen might be coded two ways: V-532 (V, 5 for N, 3 for D, 2 for S) or D-250 (D, 2 for the S, 5 for the N, 0 added).
4. Consonant separators. If a vowel (A, E, I, O, U) separates two consonants that have the same Soundex code, the consonant to the right of the vowel is coded. Example: Tymczak is coded as T-522 (T, 5 for the M, 2 for the C, Z ignored (see "side-by-side" rule above), 2 for the K). Since the vowel "A" separates the Z and K, the K is coded.

   If "H" or "W" separate two consonants that have the same Soundex code, the consonant to the right of the vowel is not coded. Example: Ashcraft is coded A-261 (A, 2 for the S, C ignored, 6 for the R, 1 for the F). It is not coded A-226.

There are other nuances: persons in religious orders were usually coded as if Brother, Sister, Father, or Mother were their surnames. Native American and Hawaiian names may be treated as in one part were a surname, such as Running Fox under F-200 for the "surname" Fox.

From "The Soundex Indexing System" http://www.archives.gov/research/census/Soundex.html.

alphabetically different but equivalent under Soundex rules. Other middle consonant substitutions are phonetically different under Soundex rules, such as Moskowitz (M-232) and Moskovitz (M-213), Potter (P-360) and Porter (P-636), and Cerwin (C-650) and Servin (S-615).

Middle consonant dropping

A name like Westurn (W-236) may be recorded instead as Weston (W-235), so the dropped "r" makes a big difference under both alphabetical and Soundex searching. Same for Stevenson (S-315) recorded as Steveson (S-312). Blake and Black (both S-420) and Presley and Pursley (both P-240) are equivalent under Soundex but quite different alphabetically.

End consonant substitution

McCullar (M-246) and McCullough (M-242) have been recorded interchangeably for the same individual or family. Another example of end consonant substitution is found in Chernow (C-650), Chernov (C-651), and Chernoff (C-651).

End consonant dropping

A name like Millard (M-463) might be misheard and recorded as Miller (M-460). When spoken, orally dropping the final "r" in McCullar and the silent "gh" in McCullough (M-242) makes both sound like McCulla (M-240), an example of end consonant dropping. Other examples are Simms (S-520) as Simm (S-500), Parker (P-626) as Parke (P-620), and McCuller (M-246) as McCalla or McAlley (both M-240). While there is no Soundex difference between Prechtel (P-623) and Precht (P-623), the extra two letters make a difference in alphabetical indexes.

End consonant addition

An extra consonant at the end of a surname is also possible, such as Miller (M-460) recorded as Millard (M-463), or Simms (S-520) recorded for Simm (S-500).

Syllable dropping

In extreme cases entire syllables might be dropped from names, such as Bush (B-200) for Terbush (T-612), and Carter (C-636) for McCarter (M-263). The opposite possibility, syllable addition, may also occur, in which James M. Carter might be misheard and misrecorded as James McCarter.

Conclusion

When we "can't find" someone in census and other historic records where we expect to find them, we need strategies to overcome the problem. Thinking creatively about possible alternative spellings for surnames (and sometimes first names) is imperative. Using the Soundex coding system as an aid to developing possible alternative spellings is a useful strategy. "B-230 W-220 F-600 M-200 S-220 I-500 T-525 I-500 S-532"![5]

Notes

1. Claire Prechtel-Kluskens, "Who Talked to the Census Taker? [1790–1870]," *NGS NewsMagazine* 31 (October–December 2005): 32–35.

2. Harold McClendon, "What Does *l?nd?n Stand For?" *Mount Vernon Genealogical Society Newsletter* 18 (January 2011): 8–9.

3. Discussions on the history of the Soundex system include Willis I. Else, *The Complete Soundex Guide* (Apollo, Penn.: Closson Press, 2002) and Tony Burroughs, "The Original Soundex Instructions," *National Genealogical Society Quarterly* 89 (December 2001): 287–98. Alternative phonetic systems have been proposed, including the New York State Identification and Intelligence system, Lawrence Phillips' Metaphone and Double Metaphone algorithms, the Daitch-Mokotoff Soundex (see Gary Mokotoff, "Soundexing and Genealogy," online http://www.avotaynu.com/Soundex.htm), and one for Indian (Hindi) languages (see Santhosh Thottingal, "Phonetic Comparison Algorithm for Indian Languages," *Santhosh Thottingal* blog, entry for 26 July 2009 online http://thottingal.in/blog/2009/07/26/indicSoundex/).

4. For Cerwin, see Charles Stephenson, "The Methodology of Historical Census Record Linkage: A User's Guide to the Soundex," *Prologue: Quarterly of the National Archives and Records Administration* 12 (Fall 1980): 151–53. For Zavodny, see Donna Pointkouski, "Census Records: What's in a Name?" *What's Past is Prologue: A Genealogy Blog*, entry for 4 February 2008 http://pastprologue.wordpress.com/2008/02/04/census-records-whats-in-a-name/.

5. Best wishes for much success in thinking in Soundex!

***Claire Prechtel-Kluskens**, a microfilm projects archivist at NARA, served as NGS registrar (1996–98) and director (1998–2000). She can be reached at ckluskens@verizon.net.*