

यूनीकोड से फोनीकोड : श्रुति क्रांति की ओर

ओम विकास

विषय बोधक शब्द : यूनीकोड, फोनीकोड, श्रुति क्रांति, पाणिनी टेबल

भारतीय भाषाएं ध्वन्यात्मक (Phonetic) हैं। सभी भारतीय भाषाओं के लिए वर्णक्रम साम्य ISCI राष्ट्रीय कोड बनाया गया है। इस 8-बिट कोड में अंग्रेजी और एक भारतीय भाषा समाहित थी और भाषा बदल ALT कुंजी से। लगभग 15 बड़ी IT कंपनियों के संयुक्त प्रयास से विश्व भाषाओं के लिए 16 बिट यूनीकोड कोड बनाया गया। इसमें भाषा-बदल की जरूरत ही नहीं। लेकिन यह लिपि पर आधारित है, जो-जो रूपिम आकृतियां अलग से प्रयोग में आती हैं उन्हें कोड में जगह दी गई है। 1991 में 9 कंपनियों ने यूनीकोड कंसोर्शियम बनाया। 2009 में 11 पूर्ण सदस्य, 4 संस्थागत, 27 एसोशिएट सदस्य हैं। भारत सरकार का सूचना प्रौद्योगिकी विभाग 2000 में सदस्य बना।

यूनीकोड (UNICODE) में IPA (इंटरनेशन फोनेटिक एल्फाबेट) के संप्रतीक अलग-अलग पेजों में है, किसी विशेष क्रम में नहीं है। पर्याप्त भी नहीं हैं। इसलिए प्रस्तावित है फोनीकोड (Phonicode)। इसकी मूल प्रेरणा ध्वन्यात्मक नागरी लिपि से मिली। “जैसा सुनो वैसा लिखो”, “जैसा लिखा वैसा बोला” सिद्धांत नागरी लेखन में है। ध्वनि उच्चारण में उच्चारण का स्थान और उच्चारण की विधि मुख्य हैं। उच्चारण के स्थान (Place of articulation) (P) के अनुसार कंठ्य, तालव्य, मूर्धन्य, दन्त्य, ओष्ठ्य 5 ध्वनि प्रकार माने गए हैं।

उच्चारण की विधि (Manner of articulation) (M) के अनुसार गले से वायु और विवर फैलाव के कम-अधिक होने के आधार पर अथवा नाक से वायु-निसरण के अनुसार 6 ध्वनि प्रकार माने गए- [(अल्प प्राण - अघोष) / (अप्र - अघ)], [(महाप्राण - अघोष) / (मप्र - अघ)], [(अल्पप्राण - घोष) / (अप्र - घ)], [(महाप्राण - घोष) / मप्र - घ] , नासिक्य, अलिजिह्वा। इन ध्वनियों को व्यंजन कहा गया। व्यंजनों को स्वर ध्वनियों से मोड्युलेट कर सकते हैं। उच्चारण स्थान के आधार पर 5 प्रकार की स्वर ध्वनियां हैं। स्वर ध्वनियां ह्रस्व,

दीर्घ हो सकती हैं। स्वर ध्वनियों परस्पर योग से व्युत्पन्न स्वर ध्वनियां 4 + 4 बनेंगी। ‘अ’ की स्वर ध्वनि को आदि स्वर और ‘उ’ की स्वर ध्वनि को मध्य स्वर और ‘म’ की व्यंजन ध्वनि को व्यंजनांत मानें तो (अ - उ - म) ध्वनि संयोग सभी ध्वनियों का द्योतक है, अर्थगत होने पर सभी संकल्पनाओं (Concepts) का। पाणिनी ने स्वर और व्यंजन समूहों को उच्चारण स्थान और विधि के अनुसार अलग-अलग वर्गीकृत किया है। ध्वन्यात्मक नागरी वर्णमाला को पाणिनी सारणी (Panini Table) में दिखा सकते हैं। जिस प्रकार रसायन विज्ञान में मेंडलीफ टेबल परमाणु क्रम को दर्शाती है जिससे रासायनिक यौगिक क्रियाओं को समझना आसान होता है, उसी प्रकार ध्वनि यौगिकों को पाणिनी टेबल से समझना आसान होगा। वर्णिकरण से अन्य संभावित ध्वनियों को जोड़ना भी आसान है।

व्यंजन के तीन उच्चारण भेद मान लें जिनमें अंतर समझना मानव कानों से तीन संभव हैं। इसी प्रकार मूल स्वर और व्युत्पन्न स्वर (ए, ओ) के तीन उच्चारण भेद अल्प, ह्रस्व, दीर्घ मान सकते हैं, इन्हें सुनकर अंतर समझ सकते हैं। संयुक्त ‘स्वर + अ’ व्युत्पन्न ध्वनियां व्यंजन की भांति हैं। इस प्रकार पहचाने जाने वाली स्वनिम/अक्षर ध्वनियों को नागरी लिपि के रूपिमों से प्रदर्शित कर सकते हैं।

अल्प नासिक्य ध्वनि को अनुस्वार ँ का रूपिम और अल्प अलिजिह्वा ध्वनि को विसर्ग ः का रूपिम दिया गया है। इन्हें स्वर श्रेणी में रखा गया है। स्वनिम - रूपिम का 1:1 (एक प्रति एक) निरूपण नागरी लिपि की विशेषता है। स्वर-स्वर, व्यंजन-स्वर, व्यंजन-व्यंजन - स्वर संयोजन से जो स्वतंत्र ध्वनियां व्युत्पन्न होती हैं, उन्हें अक्षर (Syllable) कहते हैं। अक्षर का अंत किसी स्वर से होता है। यह स्वर तीन रूपों में हो सकता — स्वर, (स्वर-अनुस्वार), (स्वर-विसर्ग)। हलन्त को व्यंजन के साथ स्वर हीन योग मान सकते हैं। अक्षरांत को नए रूपिम मात्रा से दिखाते हैं। व्यंजन को हलन्त के साथ दिखाते हैं, जैसे –

वर्ण = < स्वर, व्यंजन >
अक्षर = < स्वर, व्यंजन * स्वर >

डॉ ओम विकास, C-15, तरंग अपार्टमेंट, 19 आई. पी. एक्सटेंशन, दिल्ली - 110092

ई-मेल - dr.omvikas@gmail.com | मोबाइल - +91-9868404129

पाणिनी सारणी (Panini Table)

P = (P1,P2,P3,P4,P5), M = (M1,M2,M3,M4,M5,M6)

		व्यंजन						स्वर					स्वरांत
		अप्र- अघ	मप्र- अघ	अप्र- घ	मप्र- घ	नासिक्य	अलि जिह्वा	व्युत्पन्न	व्युत्पन्न	व्युत्पन्न	मूल	मूल	
		M1	M2	M3	M4	M5	M6	व्यंजन स्वर	दीर्घ	ह्रस्व	ह्रस्व	दीर्घ	मात्रा
कंठ	P1	क	ख	ग	घ	ङ	ह	-	-	-	अ	आ	- ा
तालु	P2	च	छ	ज	झ	ञ	श	य (इ+अ)	ऐ	ए (अ+इ)	इ	ई	ि ि
मूर्ध	P3	ट	ठ	ड	ढ	ण	ष	र (ऋ+अ)	-	-	ऋ	ॠ	ृ ॠ
दंत	P4	त	थ	द	ध	न	स	ल (लृ+अ)	-	-	लृ	लृ	--
ओष्ठ	P5	प	फ	ब	भ	म	-	व (उ+अ)	औ	ओ (अ+उ)	उ	ऊ	ू ू ो ौ

A* = <A/AA/AAA/...> पुनरावृत्ति दर्शाता है
 स्वर = < मूल स्वर (अल्प, ह्रस्व, दीर्घ), व्युत्पन्न स्वर (अल्प, ह्रस्व, दीर्घ) >
 अकारांत व्यंजन = <कम,ह, य....व>
 संयुक्ताक्षर = <क्व, क्ष, ज्ञ, प्र, श्र>
 स्वरांत/मात्रा = < ा , ि ... ु ौ ं ः >
 (ह्रस्व) मूल स्वर = < अ, इ, उ, ऋ, लृ >
 अल्प स्वर = < अँ, एँ, औँ >
 दीर्घ स्वर = < आ, ई, ऊ, ए, आ >
 व्युत्पन्न स्वर = < ए, ऐ, ओ, औ >

ध्वनियों को और गहराई से भेद करके निरूपण हेतु अलग-अलग रूपिम संप्रतीक भी दिए जा सकते हैं। संस्कृत में यह भेद किया जाता है, जिससे उच्चारण शुद्ध रहे।

इन ध्वनियों को कोडित करने के लिए व्यंजनों के लिए
 $(25+4+4) \times 3 = 87 \Rightarrow 90$ कोड प्वाइंट
 स्वरों के लिए

$(5+1) \times 3 + 2 = 20$ कोड प्वाइंट

कोडन सुविधा एवं व्यावहारिकता की दृष्टि से व्यंजन भेद दो प्रकार के ही लेते हैं, तो

(व्यंजन-स्वर) अक्षर = $(29 \times 2) \times 20 = 1160$

(द्वि व्यंजन-स्वर) अक्षर = $2 \times [(29 \times 29) \times 20] = 33640$

इनमें से लगभग 10% अर्थात् 3400 व्यावहारिक होंगे।

(त्रि व्यंजन-स्वर) अक्षर = $(29 \times 29 \times 29) \times 20 = 487780$

इनमें से लगभग 0.1% अर्थात् 500 व्यावहारिक होंगे = 500

इस प्रकार $1160 + 90 + 20 = 1270$ सामान्य अक्षर और $3400 + 500 = 3900$ संयुक्ताक्षर ध्वनियों और लगभग 500 भावात्मक (emotional) विशिष्ट ध्वनियों को भी फोनीकोड में कोडित कर सकते हैं। विशिष्ट संप्रतीकों को फोनीकोड में देख सकते हैं। फोनीकोड से यूनिकोड में परिवर्तित कर टैक्स्ट प्रिंट कर सकते हैं। स्पीच टू टैक्स्ट। यूनिकोड से फोनीकोड में परिवर्तित करके टैक्स्ट से स्पीच तैयार कर सकते हैं। फोनीकोड के हर कोड प्वाइंट के लिए स्पीच वेब फार्म संग्रहीत रहेगी। इस प्रकार नेचुरल (स्वाभाविक) स्पीच बनाने में आसानी होगी। स्वर विज्ञानी, भाषाविद् और कंप्यूटर स्पीच विशेषज्ञ की टीम फोनीकोड का मानक तैयार कर सकते हैं। भारतीय भाषाओं के लिए यह बहुत उपयोगी और आसान होगा।

स्पीच से स्पीच अनुवाद में फोनीकोड का प्रयोग उपयोगी होगा। फोनीकोड से यूनिकोड में बिना कन्वर्ट किए भी कंटेन स्टोर कर सकते हैं, इसमें मेमोरी भी कम लगेगी। 16-बिट कोड सभी भाषाओं के लिए उपयुक्त होगा। अक्षर (syllable) कोडन होने से मेमोरी की आवश्यकता अनुमानतः एक तिहाई हो जाएगी। प्रोसेसिंग स्पीड भी बढ़ जाएगी। 19 वीं सदी में औद्योगिक क्रांति में मशीनीकरण से उत्पादकता बढ़ी। 1930 के दशक से इलेक्ट्रॉनिक संचार और 1940 के दशक से कंप्यूटर के आविष्कार से सूचना क्रांति का सूत्रपात हुआ। मानव - मशीन के बीच संवाद मुख्यतः की-बोर्ड के माध्यम से होता रहा है। दो दशकों से प्रयास किए जा रहे हैं कि मानव-बोल मशीन समझ ले। यूनिकोड से विश्वभाषाओं को एक साथ रखा जा सका, फोनीकोड से मशीन मानव-बोल समझ सकेगी और मानव जैसा बोल सकेगी। इस प्रकार श्रुति क्रांति का आरंभ होगा।